DOCUMENT RESUME

ED 151 375 TM 006 793

AUTHOR	Filby, Nikola N.; Dishaw, Marilyn
TITLE	Development and Refinement of Reading and Mathematics Tests for Grades 2 and 5. Beginning Teacher Evaluation Study. Technical Report Series. Technical Report III-1. Continuation of Phase III A.
INSTITUTION	Far West Lab. for Educational Research and Development, San Francisco, Calif.
PUB DATE	15 Aug 75
NOTE	234p.; Appendices C and D of the original document--which contain test items--are copyrighted and therfore are not available nor included in the pagination; For related documents, see TM 006 794 and 795
AVAILABLE FROM	Commission for Teacher Preparation and Licensing, 1020 O Street, Sacramento, California 95814 (Appendices C and D, single copies, free of charge)
EDRS PRICE	MF-$0.83 HC-$12.71 Plus Postage.
DESCRIPTORS	Academic Ability; Academic Achievement; Achievement Gains; *Achievement Tests; Data Analysis; Educational Objectives; Effective Teaching; Elementary Education; *Elementary School Mathematics; Grade 2; Grade 5; Item Analysis; *Reading Tests; *Statistical Data; Student Evaluation; Teacher Evaluation; *Test Construction; Test Items; *Test Validity; Time Factors (Learning)

ABSTRACT

Achievement tests that are maximally sensitive to effective instruction in reading and mathematics for grades 2 and 5 were developed and refined. Important considerations regarding the tests' validity were: its coverage of instructional content (opportunity to learn), and its reactivity to instruction. Student ability must be minimally related to change in performance, and when the opportunity to learn is high, that factor must be maximally related to performance in order to have a test that is useful in studies of teaching effectiveness. An ability by opportunity to learn matrix was constructed for every item included in the test batteries. The data for each subtest are included, as well as correlations between the subtests. Sample test items are included in Appendix A. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED151375

U.S. DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SCOPE OF INTEREST NOTICE

The ERIC Facility has assigned this document for processing to:

TM

In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

CS
SE

FAR WEST LABORATORY
FOR EDUCATIONAL RESEARCH AND DEVELOPMENT

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Overview	1
Content Coverage	2
Reactivity to Instruction	6
DESCRIPTION OF TEST DEVELOPMENT PROCEDURES	11
Relationship to the Phase II Battery	11
Content Taxonomies	13
Item Populations	15
Pilot Test Procedures	21
Teacher Feedback	22
Consultant Review	24
PILOT TEST DATA	25
Overview of Data Analysis	25
Reactivity Analysis	26
Summary Data for Subtests	32
Teacher and Consultant Review	35
Summary of Revisions	38
PLANS FOR FURTHER REFINEMENT	46
REFERENCES	48
APPENDIX A. Summary Data for Subtests	50
APPENDIX B. Subtest Intercorrelations	205
APPENDIX C. Mathematics Items	*
APPENDIX D. Reading Items	*

* Bound under separate cover

LIST OF TABLES

Table		Page
I	Subtests Included in the Pilot Battery, Grade 2 Reading	16
II	Subtests Included in the Pilot Battery, Grade 5 Reading	17
III	Subtests Included in the Pilot Battery, Grade 2 Mathematics	18
IV	Subtests Included in the Pilot Battery, Grade 5 Mathematics	19
V	Sample Items with Item Statistics	30
VI	Sample Subtest Data in the Ability-OTL Matrix	34
VII	Subtests Included in the Revised Battery, Grade 2 Reading	42
VIII	Subtests Included in the Revised Battery, Grade 5 Reading	43
IX	Subtests Included in the Revsied Battery, Grade 2 Mathematics	44
X	Subtests Included in the Revised Battery, Grade 5 Mathematics	45

LIST OF FIGURES

Figure		Page
1	Opportunity-to-Learn Report Form	23
2	Ability--Opportunity-to-Learn Matrix	26
3	Sample Item Data by Ability and Opportunity to Learn	29
4	Summary of Revisions - Reading	40
5	Summary of Revisions - Math	41

INTRODUCTION

Overview

The goal of the Beginning Teacher Evaluation Study is to understand the relationship between instructional variables and student achievement in reading and mathematics at grades 2 and 5. An important part of the instrumentation of such a study is a set of measures of student achievement. This monograph presents a progress report in a continuing effort to develop appropriate and sensitive measures of student learning.

The research strategy underlying the study is one of comparison. Student achievement is measured at two points in time. Differential amounts of learning (gains) are then tested for association with characteristics of the instructional program. In order for this strategy to succeed, there must be gains in achievement and these gains must be the result of classroom learning. This fact seems obvious, yet there are a number of aspects of the usual standardized achievement test which may operate counter to the need for sensitive measures of classroom learning (Gall, 1973). Two important considerations are the issue of content coverage (opportunity to learn) and the issue of a measure's reactivity to instruction. Each of these issues will be considered in detail in the introduction to this report.

In an effort to deal with these issues, a test development effort is underway. The first step was to consider the curriculum content covered in grade 2 and grade 5 reading and mathematics. From a survey of curriculum materials, district and state objectives, achievement tests and consultant advice, four content by grade taxonomies were developed. Each

taxonomy represents a list of terminal objectives or skill areas commonly included in the curriculum. The taxonomy provides a basis for developing subtests and also for describing the content of lessons during classroom observation.

Once these taxonomies were prepared, a set of test items was selected, adapted, or developed for each content category. Some items from the BTES Phase II battery were included. Criterion-referenced tests provided another source of material.

The assembled battery of test items was pilot tested in April, 1975. In addition to standard item analysis, a reactivity analysis was carried out. This was done by assessing each item's relationship to general ability and opportunity to learn. General ability was measured by the Raven's Progressive Matrices test and by a vocabulary score. Opportunity to learn was measured by teacher ratings of the amount of time spent on each objective during the year. Performance on each item was plotted within an ability by opportunity-to-learn matrix. These data are summarized in section three of this report.

Pilot data, teacher feedback, and consultant review led to revisions in the population of items. The revised item battery will be further field tested in the coming school year (1975-76). Test items will be given three times during the year - in September, in December and in April. Teacher logs will provide information about opportunity to learn during the intertest period. This testing will provide essential information about item gains as a function of instruction. These data on item reactivity will be used for further revision of the test battery.

Content Coverage

Even within a single grade level and subject matter area, there is

considerable diversity in the specific curriculum objectives covered by teachers. Curriculum materials, instructional methods, student characteristics, and teacher preferences all influence the specific content covered in a given classroom. We are confronted with the problem how to compare achievement levels across classes when content coverage itself varies across classes.

The approach taken in most standardized test construction is to develop an achievement test which covers a broad spectrum of skills and objectives. Thus, if fractions are a part of the curriculum taught in some grade 5 math classes, then fractions are included on the test. In this way, each class can demonstrate its areas of achievement. But as a consequence, many students are tested over content which they have not covered.

On such a broadly inclusive test,(at least) two aspects of students performance are confounded, breadth of exposure and level of mastery. A student who has covered more of the content of the test will, on the average, have a higher score. Rosenshine (1971) reported studies in which the extent of content coverage in the classroom was found to be significantly correlated with test scores.

If curriculum goals are agreed upon, then the amount of material covered by a teacher is one aspect of effectiveness. But extent of coverage must be differentiated from the quality with which a teacher teaches those objectives he does cover. A teacher could bring all students to mastery on one objective yet show up poorly on an omnibus achievement test because that objective was only a small part of the content covered on the test.

The consequences of these different definitions of achievement were examined in the interpretation of the International Study of Achievement in Mathematics. Primary analyses were done on a broadly inclusive achievement battery. Fletcher (1971) called attention to the high correlation between topic coverage in different countries and achievement test scores. The rank order correlation between test scores and opportunity to learn was around .95. Fletcher proposed a reanalysis using an "efficiency ratio" of the number of items correct to the number of items students had an opportunity to learn. This efficiency ratio was defined more precisely by Romberg and Montgomery (1971). In essence, the efficiency ratio is a measure of a student's mastery of what he has been taught. Harnischfeger and Wiley (1975) suggest a similar approach.

The Peaker (1971) rejoinder to the Fletcher reanalysis pointed out the danger in relying on an efficiency ratio as the only achievement measure. Mastery will be greater, and the efficiency score higher, if only a few objectives are taught. For a complete picture of achievement one must look not only at the efficiency ratio but at the ratio between the number of objectives taught and the total possible tested content. This second ratio can be called an opportunity ratio.

In order to differentiate these two aspects of achievement it is necessary to deal with precisely defined subdivisions of content (component skills or curriculum objectives). In this way it is possible to compare teachers on the manner and success with which they teach a particular objective and also on the range of objectives which they cover.

It is clear that there are a number of levels at which content can be differentiated. At the broadest level of description are labels like

"reading" and "mathematics." These categories seem too broad, since they lead to the kind of slippage between test content and classroom content described above. On the other hand, content can be differentiated to such a fine degree that the resulting structure is too cumbersome to be practical in a naturalistic study of teacher effects. Knowledge of specific letter-sound correspondences or number facts would fall in this category.

The approach taken here has been to differentiate content hierarchically, down to the finest level of description which can be coordinated with teacher reports or classroom observation. For example, phonics skills in decoding for grade 2 are represented by specific subtests in consonant blends, long vowel with final e, vowel digraph, etc. One can ask a teacher whether or not he is teaching consonant blends, or how much attention is paid to them in the instructional program. Such teacher reports are one way of assessing the appropriateness of a subtest for measuring achievement in a particular classroom.

One can also code the content of instruction during classroom observation. If a lesson focuses on consonant blends then this provides the anchor point for linking instructional variables to achievement. Characteristics of a lesson on consonant blends should relate most closely to achievement in that specific subtest.

Beginning with the most precise manageable content coding system, one has the most direct match between the reality of instruction in the classroom and the measures of learning outcome used to assess achievement. With a hierarchical content structure, one can aggregate up to more general categories and systematically study the effects of attenuating the content match.

Reactivity to Instruction

A second major concern is the sensitivity with which tests reflect the learning which goes on in the classroom. Certainly the first requirement is that the content of the test match the content of instruction. But beyond this, one would like some evidence that the measures actually show change when instruction has taken place. Studies of instructional effects depend on sensitive measures of achievement, measures which are reactive to classroom instruction.

A number of conditions could occur to reduce a measure's reactivity. One concerns the nature of the content itself. Some areas of the curriculum may be relatively difficult to influence in school. Gains as a function of instruction are likely to be small. For instance, problem solving or inferential reasoning may be skills which are highly related to general ability, and/or background experience, and therefore, difficult to teach in school. Such skills may be an important part of the curriculum and therefore a necessary component of a balanced representation of content. But we should understand the consequences of using such measures to judge instructional effects. There is less likelihood of measuring growth related to instructional effects in an area of general processing ability than in an area of curriculum-specific knowledge (see Joyce, 1975; Berliner, 1975).

A second factor reducing reactivity is the indirectness of may measures of learning. Learning may have taken place, but we may be unable to detect it. Group testing, in particular, places constraints on the kind of test items which can be used. Test items frequently tap more than one area of knowledge or skill and so responses can be interpreted in a

number of ways. Familiarity with testing procedure is an example of general factor which enters into most achievement measures. A test score is unfortunately an imperfect indicator of any one aspect of classroom learning.

Where alternative item formats and testing procedures are available (or can be devised), some may be more closely related to the basic learning outcome than others. This is a construct validation problem. A thorough examination of the problem would require multi-trait multi-method testing. Some validity testing will be done in the fall of 1975 to compare group-administered and individually-administered measures (see Technical Report III-2). The focus in group-test development is on a specific criterion for validity - change correlated with classroom teaching. The sensitivity with which an item score shifts with instruction is a measure of the validity of that item for assessing instruction effects.

Prescriptive Reading Inventory (PRI), Sensitivity Index: One commerical test developer has become aware of the importance of reactivity to instruction. In the development of criterion-referenced tests in reading and mathematics, CAT/McGraw Hill has begun to use a "sensitivity to instruction" index. Potential test items for the Prescriptive Reading Inventory were given in the winter and again in the spring. Teachers were asked to fill out a questionnaire in which they indicated, for each objective covered in the test, whether or not they had taught to that objective during the ten week intertest period. Classes were divided into four groups on the basis of teacher response - those who had mastered the skill before the test period, those who had reviewed the skill, those who had received direct instruction on the skill, and those who had not been taught the skill. Data analysis involved an examination of change

scores for each of the four groups. Items which showed positive change (gain) for the review and instruction groups but not for the other groups would be preferred test items.

These sensitivity to instruction data were made available to the BTES staff. Some subtests were included in the BTES reading battery because they had shown promise in these data. Other subtests did not exhibit strong sensitivity yet had been included in the instructional program for a number of classes. Such subtests were generally retained in the final PRI battery and suggested areas which might be worth further examination, redefinition, or reformatting. Specific instances are discussed in the procedures section of this report.

Correlation with General Ability. Reactivity can also be assessed through an examination of the relationship between general ability and item scores. On surface examination one can predict that items which have a high correlation with general ability will be less likely to show change as a function of teaching than items which are relatively independent. Reasoning ability is itself a strong student characteristic which is difficult to influence. Items which are strongly associated with general ability are likely therefore to be resistant to instruction. Gains which occur will be more closely associated with student characteristics than with instructional variables.

Going beyond this general principal, there are several reasons why a given item might have a high correlation with general ability. One is related to the content coverage issue. High ability students move through the curriculum faster than low ability students. They are exposed to content which other students have not yet reached. Therefore, once

opportunity to learn is taken into account, some content areas may be appropriate achievement measures for high ability students only. This would produce a high correlation but is best dealt with as part of the overall opportunity to learn issue.

An item might also have a high correlation with general ability because of test format factors. The specific behavior, test procedure, and format chosen to test a skill or objective is usually one of several possibilities. The exact test procedure may or may not match activities familiar to a student from classroom experience. In effect, an element of transfer is involved for some students. High ability students are probably better able to make the connection between test items and relevant classroom experience. In part, this is a legitimate reflection of their superior grasp of content. In part, it is test taking skill. Good test takers are able to process the test material, understand its structure, weight the implications of response choices, and infer the expected behavior. A test which emphasizes the role of test taking skill, through ambiguity or unfamiliarity of format, reduces the role of classroom learning in determining performance.

Finally, an item might have a high correlation with general ability because it represents content which can be mastered more easily by high ability students. This is a legitimate manifestation of the underlying ability factor. The important question is one of degree. Consider an item which shows large gains in performance with minimal instruction for high ability students and small gains with considerable instruction for low ability students. Any overall gain in performance on such an item, say a class mean, is mostly a function of ability. Items which minimize

the ability discrepancy will be more likely to show instructional effects.

Ability-Opportunity-to-Learn Matrix. Perhaps the best way to examine reactivity is to consider ability together with opportunity to learn (OTL). The ideal data would be item performance scores at two points in time, together with information about content coverage between tests and about student ability. From these data one would be able to determine who received instruction, what improvement in performance occurred for different ability groups, and what ability differences exist prior to instruction (format factors). Such data will be collected for the revised battery (see Plans for Refinement). Interim data were collected as part of the pilot test reported here.

DESCRIPTION OF TEST DEVELOPMENT PROCEDURES

Relationship to the Phase II Battery

Test development in Phase III A began with a review of the student achievement test battery selected or developed for use during Phase II. Due to the critical importance of the achievement measures in a study of instructional effects, we felt it was important to reconsider in depth how achievement should be measured. Our basic concern was that the measures be appropriate and sensitive indicators of classroom learning, that they show change with instruction. From this perspective, a number of aspects of the Phase II battery were of concern.

The mathematics battery consisted primarily of two standardized achievement tests, the CAT Computation and Concepts tests. Both tests are omnibus measures of achievement and therefore subject to ambiguities of interpretation centered around content coverage. (Breadth of exposure is confounded with mastery of the material covered.) An analysis of the content covered in the tests indicated that the spring grade 2 form included many items beyond the grade 2 curriculum (for example, 16 division problems). This extended range means that opportunity to cover the content would be an important factor in test scores. It also means that there were relatively few items available to reflect teachers' success in teaching the content they did cover. For a study of teacher variables associated with achievement, this is a critical limitation.

The reading battery included a number of skill-specific subtests. This approach to test development seemed highly desirable. A few subtest within the reading battery seemed questionable:

(1) The Word Sounds subtest had a low correlation with other measures. Staff experience with ETU testing suggested that format might be an important factor. In the Word Sounds test the student goes through a list of ten words and circles any word which contains a target sound. This open-ended format allows the reader to stop responding when he feels he has made enough responses. In contrast, a test format which forces the student to respond to each item reduces this possible response bias.

(2) The "criterion-referenced" comprehension test contained a number of questionable items. The grade 2 test included passages which seemed confusing and uninteresting. Many questions seemed to depend strongly on a reader's background experience, reasoning ability, and/or test taking skills. A measure which emphasizes these background and ability factors is likely to be difficult to influence in school. Changes due to instruction will be small. It is, in our terms, probably not reactive to instruction.

This concern with reactivity prompted a consideration of what other measures of reading skill might be useful components of the reading test battery. The subscales included in Phase II Decoding test represent some of the possible component skills which could be tested. Other skills in decoding, word structure, word and sentence meaning could also be measured. If a traditional comprehension test measures a collection of skills and abilities, including general reasoning ability, it may be a measure with relatively low reactivity. It may be difficult to relate specific instructional variables to growth on such a measure. The more specific component skills may be more reactive to instruction. To explore this possibility it seemed reasonable to expand the reading battery and study

the reactivity of potential items. Tests for Phase III B could then be compiled by selecting the most reactive items.

Content Taxonomies

Both the theoretical issues of test development discussed in the introduction and the review of Phase II tests pointed to a need for taxonomies of content in each grade level and subject matter area. In order to match the content of instruction to test content one must have a language for describing content. Categories of content, objectives or skills, must be defined. The content taxonomy can then be used for several purposes. It can be used to identify the focus of instruction during a classroom lesson. It provides a way of communicating with teachers about content coverage. And it provides the pivotal link between classroom instruction and measures of achievement.

For the purposes of the BTES study, it seemed useful to begin with a fairly comprehensive and detailed taxonomy of content. The greater the specificity of description (up to a point) the closer the match-between classroom content and test content. Also, specific components from the taxonomy could be selected for detailed study (as in an ETU). Finally, beginning with a broad spectrum of skills allows for the selection of items on the basis of reactivity.

A number of sources were considered in the preparation of the content taxonomies:

(1) The observation instruments and work diaries from BTES Phase II included content categories.

(2) Common curriculum materials were examined. A survey of 70 teachers in BTES Phase III A indicated that 64% use the primary state

adopted math series published by Houghton-Mifflin. Another 16% use the Addison-Wesley series. It seemed reasonable, then, to focus attention on the content covered in these texts. In reading, a wider range of materials are used. Three reading series published by Harper Row, Macmillan, and Scott Foresman were most frequently used, but a number of other materials were also mentioned.

(3) State and district lists of objectives were studied. The State Reading Framework was particularly useful, especially for grade 2.

(4) Criterion-referenced tests measured specific objectives at each grade level. The Prescriptive Reading Inventory (PRI) sensitivity data was examined at this point. Teacher reports of content coverage indicated areas of concentration. Objectives which showed gain with instruction (sensitivity to instruction) were included in the taxonomy. Objectives for Grade 2 which appeared promising from this perspective were variant consonants, silent letters, number of syllables, use of comparative and superlative forms in context, contractions, and understanding word meaning from context. Comprehension items in the PRI inventory were designed according to a factorial structure but not analyzed to take full advantage of the design. This made interpretation of item change more difficult. It appeared, however, that greatest change with instruction occurred for items in sequence, character motivation, and main idea (especially for the Review group). Areas which received instruction but were lower in reactivity included definition of affixed forms and vowel with r. The objectives list and item data from the developmental phase of the PRI were very useful in development of the BTES Battery.

A comparison was made of the objectives covered in these sources and of the weight attached to each. Objectives were then selected for inclusion in the taxonomy on the basis of those most reflecting the classroom curriculum. These taxonomies, by grade level and subject matter area, are listed in Tables I through IV.

Item Populations

Once the taxonomy had been developed, the next step was to select, adopt, or develop test items in each category. Existing standardized norm-referenced tests suggested types of items, but the number of items for each objective was highly variable. Criterion-referenced tests included many of the objectives from the taxonomy. Again the number of items was variable. (The IPMS math test by Houghton-Mifflin had an abundance of items at low levels of difficulty; the PRI reading inventory often had only three items per objective.) But a number of items were adopted from existing tests.

A decision was made to aim for roughly ten items in each category. (This requirement is more nearly met in the revised battery. Time constraints made it difficult to achieve before pilot testing.) With ten items per category it is possible to evaluate items on their relative reactivity. An individual score on a subtest can be obtained with a low but not unacceptable degree of reliability. If group scores are the unit of analysis in Phase III B it will be possible to do item sampling from the population of items in each category.

Items were selected on the basis of outside data, where available. The PRI sensitivity data guided item selection. Data on the Phase II battery were examined. Unfortunately, year-long growth measures were

TABLE I
SUBTESTS INCLUDED IN THE PILOT BATTERY
GRADE 2 READING

I. Word Identification

D. Decoding (knowledge and use of letter-sound correspondences)
1. Single consonants in initial position
2. Consonant blends and digraphs in initial position
3. Consonant blends and digraphs in final position
4. Long vowel - final e pattern
5. Vowel digraphs
6. Vowel dipthongs
7. Vowel with r
8. Internal pattern in multisyllabic words
9. Consonant substitution with initial blends
10. Recognition of silent letters
11. Matching of variant consonant sounds

C. Context clues (use of sentence context to choose the correct word to fill a blank)
1. Choices vary in initial consonant
2. Choices vary in syntactic form class
3. Choices vary in word structure
4. Choices are pronouns

W. Word structure (recognition of parts of words)
1. Identification of root words
2. Division of compound words
3. Division of root word from affix
4. Number of syllables
5. Contractions

II. Word Meaning

V. Vocabulary (word meaning in isolation)
1. Synonyms
2. Definition of words with affixes

M. Word meaning in context (use of context to establish meaning)
1. Multimeaning words
2. Unfamiliar words
3. Pronoun reference

III. Comprehension

T. Text (understanding the meaning of prose material)

	Events	Description	Relationships	Main Idea
1. Literal - Recall				
2. Literal - Translation				
3. Inferential - Integration				
4. Inferential - Extrapolation				

A. Applied formats (information search)
1. Letters
2. Signs
3. Directions
4. Picture Interpretaion
5. Table of Contents
6. Riddles

TABLE II
SUBTESTS INCLUDED IN THE PILOT BATTERY
GRADE 5 READING

I. Word Identification

D. Decoding (knowledge and use of letter-sound correspondences)
1. Variant consonants
2. Short-Long vowels
3. Vowel with r
4. Silent letters

C. Context Clues (use of context to choose the correct word to fill a blank)
1. Choices vary in semantic appropriateness
2. Choices vary in word structure
3. Choices are pronouns

W. Word Structures (recognition of parts of words)
1. Identification of root words
2. Division of root word and affixes
3. Identification of first syllable
4. Identification of last syllable
5. Number of syllables

II. Word Meaning

V. Vocabulary (word meaning in isolation)
1. Synonyms
2. Definition of affixes or words with affixes

M. Word Meaning in context (use of context to establish meaning)
1. Multi-meaning words
2. Unfamiliar words
3. Pronoun referents
4. Figurative language

III. Comprehension

T. Text (understanding the meaning of prose material)

	Events	Description	Relationships	Main Idea
1. Literal - Recall				
2. Literal - Translation				
3. Inferential - Integration				
4. Inferential - Extrapolation				
5. Evaluation - Fact vs. Opinion				

A. Applied Formats (information search)
1. Letters
2. Signs
3. Directions
4. Rules
5. Maps
6. Graphs

IV. Study/Organization Skills

R. Reference Sources (choice of appropriate source of information)
1. Reference book (dictionary, encyclopedia, card catalog)
2. Reference source within a book (table of contents, glossary, index)

O. Organization of Information
1. Selecting the topic sentence from a group of sentences
2. Selecting details appropriate to a topic
3. Completing an outline

TABLE III
SUBTESTS INCLUDED IN THE PILOT BATTERY
GRADE 2 MATH

I. Computation

C1. Addition without regrouping
- a) Two 1-digit numbers, sums less than 10
- b) Two 1-digit numbers, sums 10-18
- c) One 2-digit and one 1-digit number
- d) Two 2-digit numbers
- e) Two 3-digit numbers

C2. Addition with regrouping
- a) One 2-digit and one 1-digit number
- b) Two 2-digit numbers

C3. Subtraction without regrouping
- a) Two 1-digit numbers
- b) 1-digit from a 2-digit number which is no greater than 18
- c) 1-digit from a 2-digit number
- d) Two 2-digit numbers
- e) Two 3-digit numbers

C4. Subtraction with regrouping
- a) 1-digit from a 2-digit number
- b) Two 2-digit numbers

C5. Multiplication
- a) Two 1-digit numbers, neither being greater than 5
- b) Two 1-digit numbers, one less than or equal to 5 and the other being greater than 5 but less than 10

C6. Speed test - Addition

C7. Speed test - Subtraction

C8. Number sentences involving equalities and inequalities

C9. Number line pictures

C10. Renaming numerals, equation form

II. Concepts

N. Number Systems/Number Theory
- N1. Numerals
- N2. Ordinals
- N3. Place value using compact or expanded notation
- N4. Fractions
  - a) Involving sets
  - b) Involving regions
- N5. Odd and even numbers

P. Properties
- P1. a) Associative principle of addition
  - b) Commutative principle of addition
  - c) Commutative principle of multiplication
  - d) Associative property with expanded notation

M. Measurement
- M1. Money
- M2. Linear Measurement
- M3. Measurement concepts
- M4. Liquid measurement
- M5. Time

G. Geometry
- G1. Geometric figures
- G2. Curves/Points

III. Application

A.1. Word problems

IV. Other (computation cross referenced with word problems)

- O1. Addition
- O2. Subtraction

TABLE IV
SUBTESTS INCLUDED IN THE PILOT BATTERY
GRADE 5 MATH

I. Computation

A. Whole Numbers
- C1. Addition and subtraction with numbers up to 5 digits, renaming as necessary
- C2. Multiplication with 1 or 2-digit multipliers and multiplicands up to 5-digits.
- C3. Division with 1 or 2-digit divisors, 1 or 2-digit quotients, with and without remainders

B. Rational Numbers
- C4. Add 2 proper fractions with like or unlike denominators and with sums less than or equal to 1
- C5. Add 2 proper fractions, having like denominations, sums being greater than 1 and expressed as mixed numeral
- C6. Add a mixed number to a mixed number or proper fraction, with like or unlike denominators.
- C7. Subtract a proper fraction from a whole number or another proper fraction, with like and unlike denominators.
- C8. Subtract a proper fraction or mixed number from a mixed number, with like or unlike denominators.
- C9. Multiply a proper fraction by a whole number or another proper fraction
- C10. Divide a proper fraction or whole number by a proper fraction

C. Decimals
- C11. Add or subtract decimal numbers involving tenths, hundreths, and thousandths

D. Number Sentences
- C12. Solve number sentences involving equalities and inequalities

E. Automaticity (speed test)
- C13. Addition, subtraction, multiplication, and division of whole numbers

II. Concepts

A. Numbers and Numeration System
- N1. Place value of decimal numbers involving tenths, hundreths, and thousandths

B. Properties
- N2. (a) associative property
  (b) commutative property
  (c) distributive property
  (d) identity elements

II. Concepts Cont'd

C. Number Theory
- N3. Prime Numbers
- N4. (a) multiples
  (b) common multiples
  (c) least common multiple
- N5. (a) factors
  (b) common factors
  (c) greatest common factors
- N6. Prime factorization

D. Metric Geometry
- G1. Compute perimeter
- G2. Recognize and compute area
- G3. Recognize volume
- G4. Compute surface area
- G5. Compute volume

E. Non-Metric Geometry
- G6. Line segments, parallel and perpendicular lines, angles
- G7. Geometric figures
  (a) plane
  (b) solid
- G8. Parts of a circle

F. Statistics
- S1. Number pairs
- S2. Line and bar graphs
- S3. Average, range and median

G. Functions
- S4. Number patterns, sequences

III. Application

A. Everyday Probmes
- A1. One step word problems
- A2. Two step word problems involving
  (a) addition and subtraction
  (b) addition and division
  (c) multiplication and division

B. Applying Specific Concepts
- A3. Word problems involving metric geometry
- A4. Word problems involving common measures
- A5. Word problems involving averages

IV. Other
- A6. Computation cross-referenced with word problems

available on only a few of the Phase II tests. Even within these tests, item revision made many items not strictly comparable. Nevertheless, data on item difficulty was informative. This was particularly true for the reading and math applications.

Diagnostic Items. Another consideration in item development was the value of diagnostic scales. In a well-written multiple-choice test item, the distractors (incorrect responses) represent systematic errors. One can understand something about the quality of a child's understanding or misunderstanding by examining the choices he makes. If a child consistently makes a particular type of response, this should be diagnostic of his difficulty.

It seems reasonable that the kinds of errors children make will change with instruction. If we can assess a child's partial knowledge or his progress from one type of error to another, perhaps less serious, error, we can have a more sensitive measure of his development. Teacher performance can then be judged not only on the basis of students' totally correct responses but also on their progress toward full mastery.

With such analysis in mind, some scales in the battery were designed to include systematic error types. For instance, the scale on decoding of vowel digraphs includes one distractor which is a single vowel and another which is an incorrect vowel combination. Each item consistently includes these two types of distractors. An error analysis would ask whether a child consistently chooses one or the other of these incorrect responses. We can also look at changes in response patterns following instruction.

General Ability Measures. In addition to the items assembled within

each category of the content taxonomy, two measures were included in the pilot battery to assess aspects of general ability. The first was the Raven (1960) Progressive Matrices Test. This test is regarded by many researchers as a good measure of nonverbal reasoning. The student is required to analyze a pattern with a missing piece and select the appropriate piece to complete the pattern. The Colored Progressive Matrices, sets A, AB, and B, were used at grade 2. The Standard Matrices, sets A, B, and C only, were used at grade 5.

The second measure was a vocabulary test. The California Achievement Test (CAT) Vocabulary subtest (Level 3, Form A) was used at grade 5. A portion of the CAT (Level 2, Form A) was used at grade 2, along with the BTES battery subscale on meaning of words with affixes. While vocabulary can be treated as a specific aspect of reading, a vocabulary measure is generally regarded as one of the best single measures of verbal ability (Butcher, 1968).

Pilot Test Procedures

The pilot tests were administered to six 2nd grade and six 5th grade classes. Testing was conducted over a period of one week at each grade level with each class receiving 1-1/2 hours of testing per day for three days. Days 1 and 2 involved the reading and math tests. The Raven's Progressive Matrices test and the vocabulary test were given on Day 3. All classes at a grade level received the same tests on Day 3. Due to the size of the test battery, the reading and math scales had to be distributed into two separate batteries. One group (3 classes) at each grade level was assigned the Math Computation and Reading Word Identification tests and the remaining group received the Math Application and Reading Comprehension tests.

Since test administration was conducted by the staff, no instructional test booklet was developed. Staff met and agreed on the directions to be followed. For grade 2 math all items were read orally and for grade 2 reading all sample problems and directions were read to the class. Directions, samples, and some word problems were read orally for grade 5 math and students were instructed to ask for assistance if they had difficulty with any of the word problems. The latter situation rose infrequently.

Teacher Feedback

On the weekend following each week of testing, a day long conference was held with participating teachers. Each teacher received a $50 honorarium for participating in the conference. Teachers responded to questions on the general format, administrative procedures, and content coverage of the tests. The discussion on content coverage encompassed the appropriateness of the objectives selected for testing as well as the appropriateness of each item within that scale. From the issue of appropriateness, attention became focused on the desirability of obtaining information on a student's (or student group's) opportunity to learn the skills of a particular objective. A preliminary form for recording opportunity to learn and difficulty in learning was reviewed and discussed. The resultant interchange of ideas was helpful in bringing the teachers into a test development frame of reference and gave us an opportunity to understand the teacher's frame of reference.

After the conferences, follow-up interviews were conducted to obtain Opportunity to Learn data for each student within a class. The form used to record this data is shown in Figure 1. Each teacher divided the class into High, Middle, and Low ability groups if that reflected their teaching

Figure 1
Opportunity-to-Learn Report Form

Objective:

How much time did you spend on this objective or skill during the current year?

	Not Covered	Covered Occasionally or in a 2-3 week unit			Empahsized continually or at length
Top Third	1	2	3	4	5
Middle Third	1	2	3	4	5
Bottom Third	1	2	3	4	5

strategy or into two ability groups if that was more appropriate. Ratings on Opportunity to Learn, on a one to five scale, were then completed for each tested objective and for each ability group. Class lists designating each student's math and reading ability group membership were compiled which allowed us to match students and OTL scores. Bilingual students were also identified for possible screening purposes. Ability in English was designated as fluent, limited, or non-english speaking.

Consultant Review

In addition to the feedback received from the participating teachers the pilot tests were also reviewed by consultants in reading and mathematics.[1] Here again attention was focused on the appropriateness of both the objectives selected and the items used to test those objectives. General testing procedures were also discussed. A summary of the teacher and consultant reactions is contained elsewhere in this report.

---

[1] Primary consultants to the project were Dr. Roger Shuy (reading) and Dr. Thomas Romberg (mathematics). Additional input came from Dr. Joanna Williams, Dr. Ed Begle, Dr. Priscilla Drum and Ms. Barbara Horvath.

PILOT TEST DATA

Overview of Data Analysis

Pilot test data were analyzed in a number of ways. The first analysis was a standard item analysis. This program generates scale statistics, including subtest mean, standard deviation, and internal consistency (Cronbach's alpha). It also generates the point biserial correlation between each item and total subtest score and between each item and total test score. In addition, it produces a frequency distribution of the number of students selecting each response alternative. This response distribution was especially useful, as it allowed a qualitative understanding of errors and suggested ways to improve the diagnostic properties of distractors.

Students were divided into groups on the basis of ability and of opportunity to learn a given objective. Subtest total scores were calculated for each group. In addition, each individual item P value (proportion of students passing the item) was calculated for each group. Patterns of differential performance with opportunity to learn provide an indication of an item's potential reactivity to instruction. This reactivity analysis is described in greater detail in the following section.

Summary data for each subtest are presented in Appendix A. Each subtest summary page includes the following: a description of the scale, a sample item, scale statistics, reactivity data, and a summary of revisions.

In addition to these analyses, correlation coefficients were

calculated between all pairs of subtest total scores for each group of subjects. Subtest total scores were also correlated with total scores on Ravens and on Vocabulary. These inter-correlations are shown in Appendix B.

Besides these numerical analyses, general qualitative feedback was obtained from teachers and consultants. A summary of their comments and suggestions is contained in the section on teacher and consultant review.

Once all these sources of input had been considered, the item populations were revised. Some scales were dropped completely, some new scales were added. Most scales were modified in some way, either through item deletion, addition, or modification. These changes are summarized in the last part of this section. A revised list of content categories is presented. Details of revision are indicated on each subtest summary page.

Reactivity Analysis

Procedure. The pilot testing included two variables which are potentially relevant to item reactivity - general ability and opportunity to learn (OTL). These two variables combine to create a two-dimensional matrix, as shown in Figure 2.

Figure 2
Ability - Opportunity-to-Learn Matrix
Opportunity to Learn

Ability	1 (no direct instruction)	2	3	4	5 (substantial direct instruction)
Low					
Middle					
High					

For any one subtest, each student can be assigned to one cell in the matrix. One can then calculate mean subtest total socres within each cell. Patterns of change in performance level across the matrix indicate the general reactivity of each objective or skill area. In addition to mean subtest toal scores, each individual item P value (proportion correct) can be calculated within each cell. Examining differences between items can help pinpoint areas of greater reactivity. Both of these procedures were carried out for each subtest.

Since two ability measures were used, one verbal (vocabulary) and one nonverbal (Ravens), two matrices of data were generated for each subtest. For each ability measure, the group of students taking a set of tests was divided into three levels of roughly equal size. By this procedure, labels like "high" and "low" ability are strictly relative. The interest was not in examining norm-based ability designations but rather in looking for differences within the available sample.

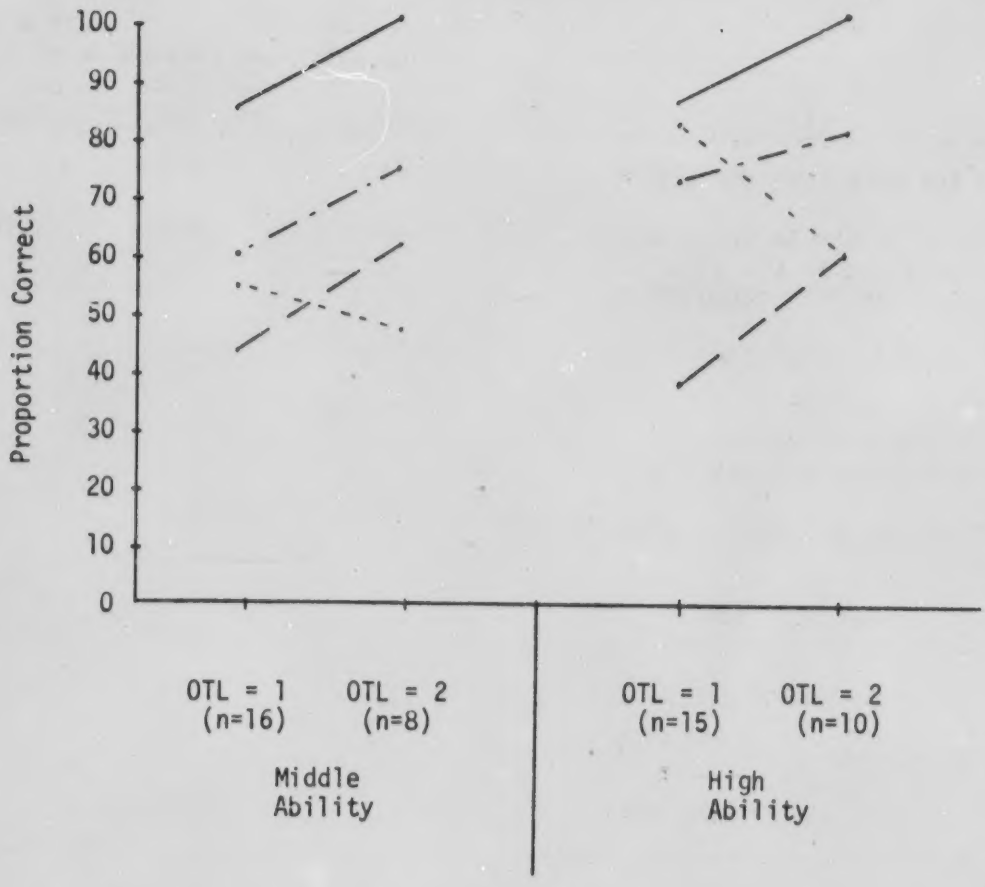
One additional complication arose. Some grade 5 students shifted to other teachers in the school for part of their reading instruction. For this reason, we could not get precise opportunity to learn ratings for all students on each reading objective. During the weekend discussion with these teachers, it appeared that some objectives were covered in "homeroom" language arts instruction. In a self-contained class a teacher could report on all aspects of instrction at one time. For several of these teachers, different students were involved for reading and for language arts. Therefore, a separate OTL rating was made for language arts instruction where appropriate. Separate matrices were generated for reading OTL's and language OTL's.

Discussion. The principal question of interest in the reactivity analysis is whether performace improves with increasing opportunity to learn. One might predict that students who have received no direct instruction in a particular skill area (OTL=1) will perform less well than students who have received instruction in the skill (OTL2-5). This would result in high P values as one moves across the opportunity to learn dimension, a positive slope if the data were presented graphically.

Data for selected sample items from subtest M1 in grade 2 are graphed in this manner in Figure 3. The items themselves, along with correlations from the item statistics analysis, are presented in Table V. An examination of Figure 3 shows that over half the students of middle and high ability defined by scores on the vocabulary test, received no direct instruction in the skill of defining multi-meaning words in the context. All of the low ability students fell in this no instruction category, so they have been omitted from the Figure. Those students who did receive instruction tended to perform better than the no instruction group. Three of the four items show positive slopes. The remaining item, item 10, shows a decline in performance for both ability groups. Moreover, item 10 shows the greatest difference between ability groups. The difficult vocabulary in item 10 is probably an important performance factor. Following our criterion of selecting reactive items, item 10 will be revised or replaced in the test battery.

It is useful to compare the information available from the reactivity matrix with the information available from standard item analysis. The question is would the same decision be reached from other test development criteria. As seen in Table y, nothing in the item correlations indicates

Figure 3
Sample Item Data by Ability and Opportunity-to-Learn



Key:

———— Item 2

-------- Item 4

-·-·-·-· Item 6

····· Item 10

TABLE V
SAMPLE ITEMS WITH ITEM STATISTICS

Item	Proportion of Students Responding Correctly	Correlation With Total Test Score	Correlation With Total Score on Subtest
2. Pour the milk into the glasses.			
A. some things to drink out of B. something to see with C. children in a schoolroom	.84	.36	.50
4. His pen was out of ink.	.38	.33	.36
A. to write B. a place for animals C. something to write with			
6. You have to duck to get into the cave.	.62	.49	.55
A. a kind of bird B. move your head down C. a place to tie a boat			
10. The pilot will land at the airport.	.53	.51	.62
A. a country B. soil; earth C. to bring an airplane down			

any negative properties to item 10. If an item were to be eliminated on the basis of correlations, it would be Item 4. Such a decision would in fact reduce the test's sensitivity to instruction. Discrepancies like this foster the hope that relying on a reactivity index for item selection will in fact produce a test which is more sensitive to instructional effects than previous instruments.

Unfortunately, not all of the subtests produced useful data on item reactivity. The essential difficulty is that only one performance measure, a spring posttest score, is available for each item. Without information about fall performance levels there is no information about change scores. The interpretation given above rests on the assumption that scores for students with no opportunity to learn are acceptable substitutes for fall performance levels. If anything, they should be a generous estimate, since general reading practice could improve performance over the year. Overestimating fall scores would reduce the likelihood of showing improvement with instruction. Positive trends which appear in spite of this limitation are an encouraging indicator.

Several factors can eliminate this fall score substitute: (1) There may not be any students with no direct instruction. This is frequently the case where a skill is an important part of all curricular approaches. Differences in performance level with different amounts of instruction might still occur, but they are likely to be a severe underestimate of reactivity. In some cases students with different amounts of instruction perform at roughly the same level, that is the lines in the graph are flat. Without pretest scores, there is no way to assess reactivity from such patterns.

(2) Students with no direct instruction may show mastery of a skill. This situation would occur if students had mastered the skill prior to grade 2 and therefore did not need instruction. Knowledge of the sequence of instruction in the curriculum allows one to identify subtests where this pattern is likely to occur. If students at all ability levels show mastery, it further supports this interpretation. If many students show mastery without instruction it suggests that an objective is too easy to be included in the battery. But if instruction does occur for a reasonably large number of students it may be a mistake to eliminate the objective at this point. Without fall scores it is impossible to conclude that all students reach mastery before grade 2. Learning may still occur for some students. Content coverage should be the important factor at this point.

Even if reactivity cannot be assessed directly from the data pattern, the ability -OTL - matrix still provides useful information. The OTL ratings alone indicate content coverage in the classroom. Content areas which recieve considerable attention during the year ought to be included in a measure of classroom learning. Considering OTL ratings along with ability indicates for whom instruction is likely to occur and for whom it is likely to pay off. Strong ability differences at OTL=1, or consistently across the matrix, suggest the need to examine format factors. Variability in pattern across items in a subtest can help identify deviant items.

Summary Data for Subtests

Data were analyzed separately for each item and in a number of separate analyses. Such analyses lead to an overabundance of detail. In reporting the data, it seemed most useful to include overall subtest data. A

summary data page for each subtest is included in Appendix A. Each summary data page includes the following elements:

(1) A description of the subtest

(2) A sample item

(3) Scale statistics

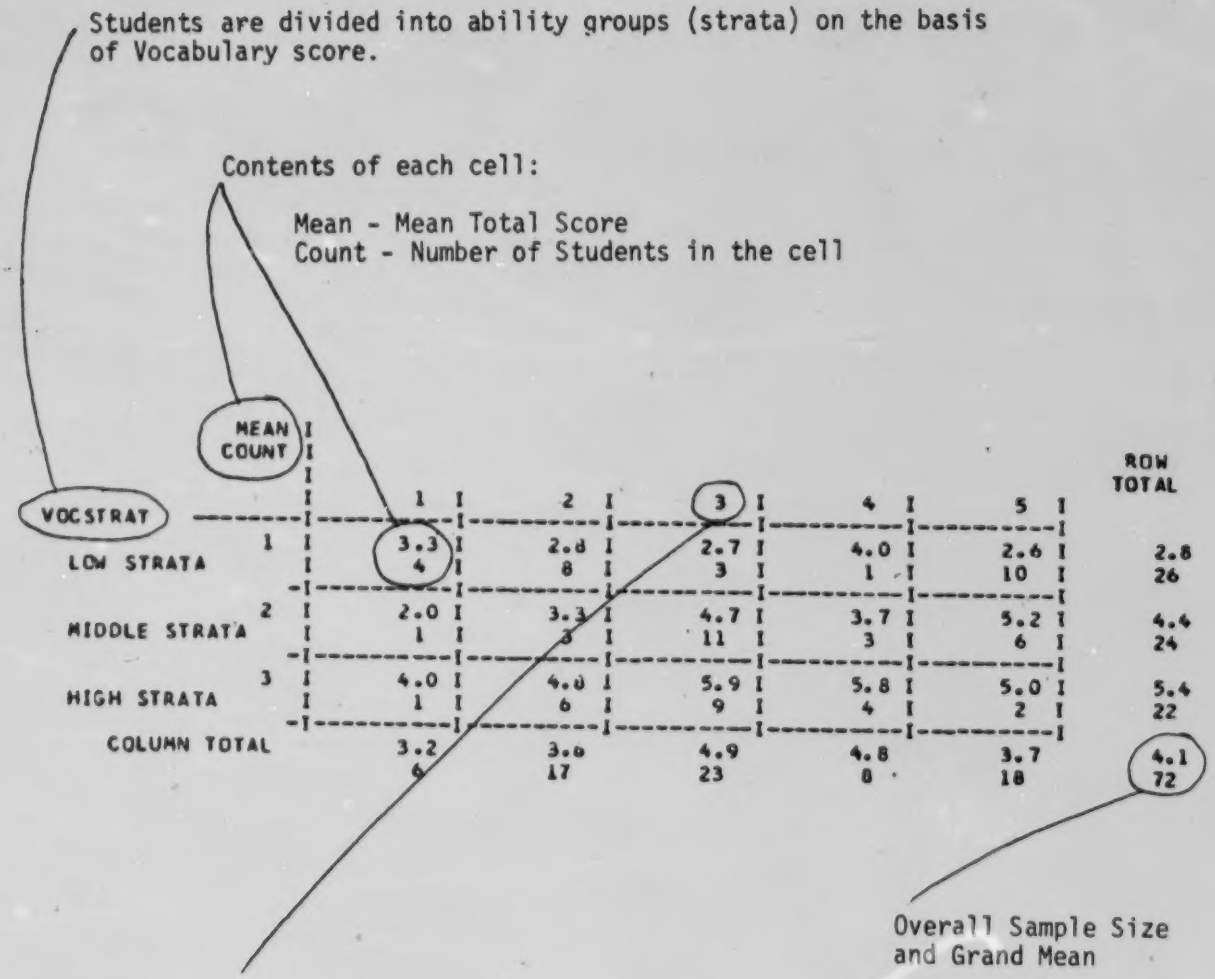
(4) A table showing mean total score and sample size within each cell of an ability by opportunity-to-learn matrix

(5) A summary of revisions made on the basis of pilot test data.

Important information about individual items appears in the discussion of revisions.

Most elements of the summary data pages are self-explanatory with the possible exception of the computer-generated ability-OTL Matrix. A sample matrix, with explanations, is in Table VI.

TABLE VI
SAMPLE ITEM DATA IN THE ABILITY-OTL MATRIX



Teacher and Consultant Review

Mathematics. The day long conferences with participating teachers were quite helpful particularly in the specificity of their comments on scale construction as related to classroom instruction.

On the grade 2 math computation scales the teachers recommended the use of alternate sentence forms which would place the box in a position other than terminal. They also recommended the inclusion of a scale on number patterns/sequences. The scales on liquid measurement and odd and even numbers were considered inappropriate, as they represent supplementary content not covered by most of the teachers. They also felt that the number line scale was testing a style of teaching which most of them did not use. Their other recommendations, aside from the need for primary type, revolved around the organization of items within a scale, the phrasing of specific questions or directions, and the possible ambiguity of some items. The majority of their recommendations were incorporated in the revised test battery. Teacher influence on revisions is notated on the data summary pages.

Consultant review of grade 2 math was in agreement with most of the teacher recommendations. However, the opinion was given that the scale on time was trivial and should be dropped or revised into a problem solving format. Discussion with the consultant also focused on the number line scale and the problem of different representational approachs. This problem was not resolved but the recommendation was given to include pictorial representation in other scales, like place value, where the problem of differing styles is not as great and generalization across styles is more likely to occur. Specific suggestions were also given on improving the items

testing numerals, fractions, money and measurement concepts. Most of these suggestions were incorporated in the revised test battery.

In the grade 5 teacher conference discussion immediately focused on the number and range of math objectives that were included in the pilot tests. Though the range of objectives was reflective of the texts being used, they were not reflective of actual classroom instruction. The focus of instruction for many pupils at this level is on the computation of whole numbers. Thus, the teachers' first recommendation was to expand the scales involving whole numbers and to simplify and reduce the emphasis on fractions. None of the participating teachers had covered any of the number theory material on multiples and very few had covered prime numbers, factors or prime factorization. This same problem was encountered with the metric and non-metric geometry objectives. Geometry appears to be treated as "elective" material and the time allotted to it's teaching varied considerably. Unfortunately, those classes which received the most instruction on geometry were not among the classes which were tested on these objectives. Revision comments on the summary data pages indicate the extent to which these scales were modified due to teacher feedback. As in the grade 2 conference, attention was also given to the organization of items within a scale, the phrasing of questions, and the identification of particular items which they felt were inappropriate for a variety of reasons.

Consultant review supported the teacher's criticism that some of the test material was more appropriate for 6th grade. Specific suggestions on item revisions were given and subsequently incorporated into the revised battery.

Reading. The Saturday conference with grade 2 teachers was an informative session. Discussion began with the format of the decoding tests. Teachers raised concerns about the use of nonsense words, and more generally about the use of whole "words" rather than letters. There was general discussion about the problems involved in group-testing formats where individual testing would allow a more direct assessment. The reasons for a whole "word" format were explained (more like context of reading; indicates the positon of the target sound). The explanation and discussion aleviated some of the concern.

In the context clues section, teachers reported considerable classroom instruction. They indicated that little attention was given to possessives and pronouns or to the f-ves spelling change.

Teachers reported much use of the context clues format, and relatively little use of the word meaning in context (underlined word) format.

Consultant feedback covered some of the same areas outlined above. Again there was a discussion of nonsense words, although the value of using whole "words" rather than letters was clear to the consultant. Particular likely confusions in letter sounds were pointed out and used to improve distractors. The consultant was critical of the "silent letters" concept. He pointed out that some series treat combinations like gn as letter pairs, rather than a silent letter, and indicated that linguistically the phenomenon is not as simple as is often presented. Differences in the way the concept is presented in classrooms make it difficult to test fairly.

The consultant particularly liked the word structure sections on word division (W2 and W3), and the section on word meaning in context. In the comprehension sections, he stressed the importance of understanding the

main idea.

Grade 5 teachers generally felt that decoding was a small part of reading instruction, except for poor readers. They indicated, though, that other aspects of instruction, particularly spelling and grammar, did touch on these skills. In effect, the teaching of language-related skills is more dispersed at grade 5. For this reason, teachers gave language arts OTL scores for some objectives.

In regard to comprehension, a general discussion was held about kinds of tasks used to assess and/or improve comprehension. The categories of comprehension in the matrix were somewhat unfamiliar, but with some discussion and examples became reasonably clear. Teachers were encouraged to give their own ideas about how to describe comprehension. One point that emerged is that evaluation of facts and opinions is largely a supplementary activity, studied only with good readers. Outlining was not covered by any of the teachers. Some of the applied comprehension items looked interesting to the teachers, but in general little time was spent on non-text material.

The consultant called attention to differences in syllabification rules used by different programs, a potential source of ambiguity in interpreting responses on scales W3 and W4. He also pointed out a number of ambiguous items in scale D2, decoding of long and short vowels.

Summary of Revisions

On the basis of the pilot testing, a number of revisions were made in the test batteries. Major revisions are listed in Figures 4 and 5. Minor revisions are indicated on the individual data summary pages. The list of subtests in the revised test battery for each grade level and

subject matter area is given in Tables VII through X. The tests themselves are located in appendices detached from this report. Appendix C contains the mathematics items. Appendix D contains the reading items. These appendices contain subtest pages placed in order and identified according to Tables VII through X.

Measures of Attitude. In addition to the cognitive measures discussed in this report, measures of attitude have been added to the battery. We have chosen items which are as direct as possible and which may be administered straightforwardly in a short period of time. The same items are being used at grades 2 and 5 with slightly different response formats. In both cases students make a mark along a line indicating the degree to which they like the concept in the stem of the item. At grade 2, students mark along a line between a smiling face and a frowning face; at grade 5 the line is numbered and the end points are labeled "like a lot" and "do not like". There are two items which relate to reading and two which relate to mathematics. Scores will be added over the items within content areas. Each set of items begins with a practice item about ice cream. In the main, responses to this item are expected to be highly positive and have relatively small variance. The distribution of responses on this item wll provide some information on whether or not students have understood the task. The last item in the set asks directly how much students like school. Although these items are brief and direct, it is expected that they will provide indications of the direction of change in attitude over the year. These attitude items are included at the end of Appendices C and D.

Figure 4
Summary of Revisions - Reading

I. Grade 2

A. Scales Reorganized
1. Decoding of initial single consonants and of initial and final blends collapsed into one category-consonants. Items divided into a power test and a speed test.
2. Variant consonants test reformatted

B. Scales deleted
1. Silent letters
2. Context clues - pronouns
3. Definition of words with affixes

C. Scales substantially expanded (+) or reduced (-)
1. Long vowel - final e +
2. Digraphs +
3. Contractions +
4. Pronoun reference +

D. Scales added
1. Sight words
2. Comprehension of text with picture responses

II. Grade 5

A. Scales deleted
1. Silent letters
2. Application - graphs
3. Outlines

B. Scales substantially expanded (+) or reduced (-)
1. Variant consonants -
2. Short-Long vowels -
3. Vowel with r -
4. Signs -

Figure 5
Summary of Revisions - Math

I. Grade 2

A. Scales deleted
1. Number line pictures
2. Odd and even numbers
3. Liquid measurement
4. Time

B. Scales substantially expanded (+) or reduced (-)
1. Place value +
2. Measurement concepts -
3. Fractions +

C. Scales added
1. Number patterns

II. Grade 5

A. Scales deleted

1. Multiplication of fractions
2. Division of fractions
3. Number sentences involving equalities and inequalities
4. Prime numbers
5. Multiples
6. Parts of a circle
7. Word problems involving metric geometry
8. Statistics: average/range/median

B. Scales substantially expanded (+) or reduced (-)
1. Addition of whole numbers +
2. Subtraction of whole numbers +
3. Multiplication of whole numbers +
4. Division of whole numbers +
5. Addition of fractions -
6. Subtraction of fractions -
7. Factors -
8. Area (recognize & compute) -
9. Volume (recognize & compute) -
10. Geometric figures -
11. Line and bar graphs +

C. Scales added
1. Equivalent fractions

TABLE VII
SUBTESTS INCLUDED IN THE REVISED BATTERY
GRADE 2 READING

I. Word Identification

D. Decoding (knowledge and use of letter-sound correspondences)
1. Consonant sounds (power test)
2. Consonant sounds (speed test)
3. Variant consonant
4. Long vowel - final e pattern
5. Vowel digraphs
6. Vowel dipthongs
7. Vowel with r
8. Internal pattern in multisyllabic words
9. Consonant substitution with initial blends
10. Sight words

C. Context clues (use of sentence context to choose the correct word to fill a blank)
1. Choices vary in initial consonant
2. Choices vary in syntactic form class
3. Choices vary in word structure

W. Word structure (recognition of parts of words)
1. Identification of root words
2. Division of compound words
3. Division of root word from affix
4. Number of syllables
5. Contractions

II. Word Meaning

V. Vocabulary (word meaning in isolation)
1. Synonyms (California Achievement Test)

M. Word meaning in context (use of context to establish meaning)
1. Multimeaning words
2. Unfamiliar words
3. Pronoun reference

III. Comprehension

T. Text (understanding the meaning of prose material)

	Events	Description	Relationship	Main Idea
1. Literal - Recall				
2. Literal - Translation				
3. Inferential - Integration				
4. Inferential - Extrapolation				
5. Text with picture responses				

A. Applied formats (information search)
1. Letters
2. Signs
3. Directions
4. Picture Interpretaion
5. Table of Contents
6. Riddles

TABLE VIII
SUBTESTS INCLUDED IN THE REVISED BATTERY
GRADE 5 READING

I. Word Identification

D. Decoding (knowledge and use of letter-sound correspondences)
1. Variant consonants
2. Short-Long vowels
3. Vowel with r

C. Context Clues (use of context to choose the correct word to fill a blank)
1. Choices vary in semantic appropriateness
2. Choices vary in word structure
3. Choices are pronouns

W. Word Structure (recognition of parts of words)
1. Identification of root words
2. Division of root word and affixes
3. Identification of first syllable
4. Identification of last syllable
5. Number of syllables

II. Word Meaning

V. Vocabulary (word meaning in isolation)
1. Synonyms (California Achievement Test)
2. Definition of affixes or words with affixes

M. Word Meaning in context (use of context to establish meaning)
1. Multi-meaning words
2. Unfamiliar words
3. Pronoun referents
4. Figurative language

III. Comprehension

T. Text (understanding the meaning of prose material)

	Events	Description	Relationships	Main Idea
1. Literal - Recall				
2. Literal - Translation				
3. Inferential - Integration				
4. Inferential - Extrapolation				
5. Evaluation - Fact vs. Opinion				

A. Applied Formats (information search)
1. Letters
2. Signs
3. Directions
4. Rules
5. Maps

IV. Study/Organization Skills

R. Reference Sources (choice of appropriate source of information)
1. Reference book (dictionary, encyclopedia, card catalog)
2. Reference source within a book (table of contents, glossary, index)

O. Organization of Information
1. Selecting the topic sentence from a group of sentences
2. Selecting details appropriate to a topic

TABLE IX
SUBTESTS INCLUDED IN THE REVISED TEST BATTERY
GRADE 2 MATH

I. Computation

C1. Addition without regrouping
- a) Two 1-digit numbers, sums less than 10
- b) Two 1-digit numbers, sums 10-18
- c) One 2-digit and one 1-digit number
- d) Two 2-digit numbers
- e) Two 3-digit numbers

C2. Addition with regrouping
- a) One 2-digit and one 1-digit number
- b) Two 2-digit numbers
- c) Three 1 or 2-digit numbers

C3. Subtraction without regrouping
- a) Two 1-digit numbers
- b) 1-digit from a 2-digit number which is no greater than 18
- c) 1-digit from a 2-digit number
- d) Two 2-digit numbers
- e) Two 3-digit numbers

C4. Subtraction with regrouping
- a) 1-digit from a 2-digit number
- b) Two 2-digit numbers

C5. Multiplication
- a) Two 1-digit numbers, neither being greater than 5
- b) Two 1-digit numbers, one less than or equal to 5 and the other greater than 5 but less than 10

C6. Speed test - Addition

C7. Speed test - Subtraction

C8. Number sentences involving equalities and inequalities

C9. Renaming numerals equation form

II. Concepts

N. Number Systems
- N1. Numerals and Ordinals
- N2. Place value with compact or expanded notation
- N3. Fractions
  - a) Involving sets
  - b) Involving regions and lines
  - c) Quantatative relationships

P. Properties
- P1. a) Associative principle of addition
  - b) Commutative principle of addition
  - c) Commutative principle of multiplication
  - d) Identity elements of addition and subtraction
- P2. Associative property with expanded notation

M. Measurement
- M1. Money
- M2. Linear measurement
- M3. Measurement concepts

G. Geometry
- G1. Geometric figures
- G2. Curves and points

III. Application

A1. Word Problems

IV. Other

O1. Number patterns/sequences
O2. Addition (word problem-cross referenced)
O3. Subtraction (word problem-cross referenced)

TABLE X
SUBTESTS INCLUDED IN THE REVISED TEST BATTERY
GRADE 5 MATH

I. Computation

A. Whole Numbers
C1. Addition with numbers up to 5-digits, renaming as necessary
C2. Subtraction with numbers up to 5-digits, renaming as necessary
C3. Multiplication with 1 or 2-digit multipliers and multiplicands up to 5-digits
C4. Division with 1 or 2-digit divisors, 1 to 3-digit quotients, with and without remainders

B. Rational Numbers
C5. Identify equivalent fractions
C6. Addition of
(a) 2 proper fractions with like or unlike denominators, and with sums less than or equal to 1
(b) 2 proper fractions with like denominators, sums being greater than 1 and expressed as a mixed numeral
(c) a mixed number to a mixed number or proper fraction, like or unlike denominators
C7. Subtraction of
(a) a proper fraction from a whole number or another proper fraction, like or unlike denominators
(b) a proper fraction or mixed number from a mixed number, with like or unlike denominators

C. Decimals
C8. Add or subtract decimal numbers involving tenths and hundredths

D. Automaticity (speed test)
C9. Addition, subtraction, multiplication, and division of whole numbers

II. Concepts

A. Numbers and Numeration System
N1. Place value of
(a) decimals - involving tenths and hundredths
(b) whole numbers thru to millions

B. Properties
N2. (a) associative property
(b) commutative property
(c) distributive property
(d) identity elements
(e) inverse relationships

II. Concepts Cont'd

C. Number Theory
N3. Factors and prime factorization

D. Metric Geometry
G1. Recognize and compute
(a) perimeter
(b) area
(c) volume

E. Non-metric Geometry
G2. Recognize and identify
(a) line segments, parallel and perpendicular lines, lines, angles
(b) plane geometric figures

F. Statistics
S1. (a) interpret line and bar graphs
(b) identify number pairs on number plane

G. Functions
S2. Number patterns and function rules

III. Application

A1. One step word problems

A2. Two step word problems
(a) any two basic processes
(b) averages

A3. Word problem involving common measures

IV. Other

A4. Computation cross-referenced with word problems.

PLANS FOR FURTHER REFINEMENT

The items resulting from this year's developmental effort will be field tested during the coming school year. A discussion of plans for test refinement can be found in the Program Plan for Continuation of Phase III A. Essentially, the plan is that the battery of test items will be given three times during the year - in September, in January, and in April. These repeated testings will allow us to study changes in performance on each item over the course of the year. At the same time, teachers will keep records of the content covered in classroom instrction. These teacher logs are coordinated with Task II, a study of instructional time and its relationship to achievement. A description of teacher report procedures can be found in Technical Notes II-1 and II-2.

In order to carry out the field testing, the subtests in Appendices C and D will be organized into administerable test batteries. Due to the large number of subtests two test batteries will be assembled for each subject matter and grade level (as was done this year for pilot testing). Subtest units will be kept intact, and each subtest will be assigned to one of the two groups. Final decisions about assignment to groups will depend on coordination with Task II requirements. The overall design for test administration is the following:

	Day 1		Day 2	
	Math	Reading	Math	Reading
Group A	45 min.	45 min.	45 min.	45 min.
Group B	45 min.	45 min.	45 min.	45 min.

As decisions are made about the order of subtests for administration, it will probably be necessary to add general sample problems or directions. A manual for test administration will also be prepared. All grade 2 test pages will be enlarged when they are printed.

Fifteen teachers at each grade level have been recruited to participate in the field testing. These teachers are drawn from several districts in the San Francisco area.

The data generated in this field test will provide a good opportunity to examine reactivity to instruction. Item change scores will be plotted within an ability-OTL Matrix. Items which show gains as a function of instruction will be prime candidates for the Phase III B test battery. It is believed that this refinement procedure will result in a more sensitive measure of achievement and will enhance the Phase III B research effort.

REFERENCES

1. Basic Objectives for California State Reading Tests for Grades 2, 3, 6, and 12. Sacramento: California State Department of Education, 1974.

2. Berliner, David. Impediments to the study of teacher effectiveness. Technical Report No. 75-6-3, Far West Laboratory, June, 1975.

3. Bloom, Benjamin S. (Ed). Taxonomy of Educational Objectives. Handbook I.: Cognitive Domain. New York: David McKay, 1956.

4. Butcher, H. J. Human Intelligence: Its nature and assessment. London: Methuen & Co. Ltd., 1968.

5. Filby, Nikola. Design for the use of individual tests. Technical Note III-2, Far West Laboratory, September, 1975.

6. Fletcher, Harold J. Critiques of the International Study of Achievement in Mathematics: An efficiency reanalysis of the results. Journal for Research in Mathematics Education, 1971, 2 (2), 143-156.

7. Gall, Meredith. The problem of "student achievement" in research on teacher effects. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, February 1973.

8. Harnischfeger, Annegret and Wiley, David. Notes for a field study of instructional time. Working Paper, Far West Laboratory, June 1975.

9. Joyce, Bruce. Variables from the study of teaching: Priorities and concerns about measurement. Technical Report No. 75-6-4, Far West Laboratory, June 1975.

10. Leiderman, Gloria. ELMA Reports: No. 1, a longitudinal study of mathematical achievement in the primary school years: Description of design, sample, and factor analyses of tests. Palo Alto, California: Stanford University, 1971.

11. Peaker, Gilbert. A rejoinder for IEA. Journal for Research in Mathematics Education, 1971, 2 (2), 158-163.

12. Program Plan for Continuation of Phase III A Beginning Teacher Evaluation Study. Far West Laboratory, May 1975.

13. Raven, J. C. Guide to the Standard Progressive Matrices. London: H. K. Lewis & Co. Ltd., 1960.

14. Romberg, Thomas and Montgomery, Mary. Comments on Fletcher's efficiency ratio. Journal for Research in Mathematics Education, 1971, 2 (3), 235-237.

15. Romberg, Thomas A. and Wilson, James. NLSMA Report No. 7. The Development of Tests. Stanford University, 1969.

16. Rosenshine, Barak. Teaching behaviours and student achievement. London: National Foundation for Educational Research in England and Wales, 1971.

17. Sipay, Edward R. Sipay Word Analysis Tests. Cambridge, Mass.: Educators Publishing Service, Inc., 1974.

APPENDIX A

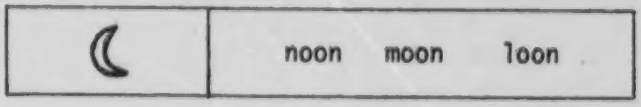
Summary Data For Subtests

	Page
Grade 2 Reading	51
Grade 5 Reading	85
Grade 2 Mathematics	124
Grade 5 Mathematics	160

D1. Decoding - Single Consonants

Tests the student's knowledge of consonant sounds. Student must select the word which names a picture. Choices vary in initial consonant, with distractors including letters which are visually and/or acoustically similar to the correct choice.

Sample:



Pilot Data:

Number of Items	10
Number of Students	69
Mean Total Score	9.64
Standard Deviation	1.55
Mean Proportion Correct	.96
Internal Consistency	.95

Ability-OTL-Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA	0.0 0	0.0 0	10.0 1	10.0 6	8.3 13	8.9 20
2 MIDDLE STRATA	10.0 1	10.0 3	10.0 9	10.0 1	10.0 6	10.0 20
3 HIGH STRATA	9.9 8	10.0 2	10.0 3	10.0 4	0.0 0	9.9 17
COLUMN TOTAL	9.9 9	10.0 5	10.0 13	10.0 11	8.8 19	9.6 57

Revisions: Some distractors were changed to make the types of distractors more consistent. This allows for an analysis of systematic error patterns. Six items from this scale were moved into the speed test. Four items were retained in the power test.

D2. Decoding - Consonant blends and digraphs in initial position

Tests ability to decode consonant blends and digraphs in initial position. Student must select the word which names a picture. Choices vary in initial consonant configuration. One distractor has only a single initial consonant; another has a blend or digraph containing one of the two letters in the correct choice.

Sample:

swing sing sting

Pilot Data:

Number of Items	12
Number of Students	69
Mean Total Score	11.00
Standard Deviation	1.69
Mean Proportion Correct	.92
Internal Consistency	.77

Ability-OTL Matrix:

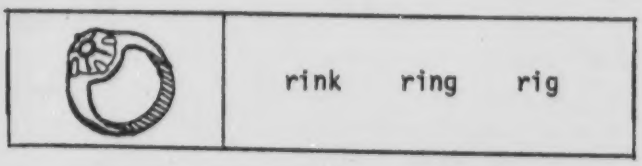
MEAN COUNT / VOCSTRAT	1	2	3	5	ROW TOTAL
1 LOW STRATA	0.0 0	11.4 10	12.0 1	8.8 9	10.3 20
2 MIDDLE STRATA	12.0 1	11.2 9	11.9 9	11.0 1	11.5 20
3 HIGH STRATA	11.6 8	11.6 8	11.0 1	0.0 0	11.6 17
COLUMN TOTAL	11.7 9	11.4 27	11.8 11	5.0 10	11.1 57

Revisions: Distractors in two items were changed so that the single letter included in all choices was the same as the first letter of the correct response (as in the sample above). Seven items were moved into the speed test.

D3. Decoding - Consonant blends and digraphs in final position

Similar in format and purpose to D2, but with choices varying in final consonant configuration.

Sample:



Pilot Data:

Number of Items	10
Number of Students	69
Mean Total Score	9.06
Standard Deviaton	1.74
Mean Proportion Correct	.91
Internal Consistency	.80

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA	0.0 0	9.6 10	10.0 1	0.0 0	7.0 9	8.4 20
2 MIDDLE STRATA	10.0 1	9.5 6	9.9 9	8.3 3	10.0 1	9.5 20
3 HIGH STRATA	9.9 8	10.0 8	10.0 1	0.0 0	0.0 0	9.9 17
COLUMN TOTAL	9.9 9	9.7 24	9.9 11	8.3 3	7.3 10	9.3 57

Revisions: Six items were moved into the speed test.

D4. Decoding - Long vowel - final e pattern

Tests student's ability to apply the generalization that in a CVCe pattern the e is silent and the vowel sound is long. The student must select the correct word to name a picture. The presence or absence of the e is the critical distinction. An additional distractor in the pilot version used a consonant blend in final position.

Sample:

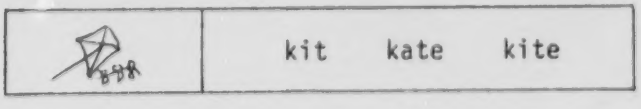
Pilot Data:

Number of items	5
Number of Students	69
Mean Total Score	4.14
Standard Deviation	1.27
Mean Proportion Correct	.83
Internal Consistency	.73

Ability-OTL Matrix:

VOCSTRAT	MEAN CCLNT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA	2.8	2.8	3.8	4.7	0.0	0.0	3.6
	9	4	7	0	0	20	
2 MIDDLE STRATA	2.0	4.8	4.6	0.0	4.7	4.5	
	1	6	10	0	3	20	
3 HIGH STRATA	0.0	4.8	5.0	5.0	0.0	4.9	
	0	8	7	2	0	17	
COLUMN TOTAL	2.7	4.6	4.8	5.0	4.7	4.3	
	10	18	24	2	3	57	

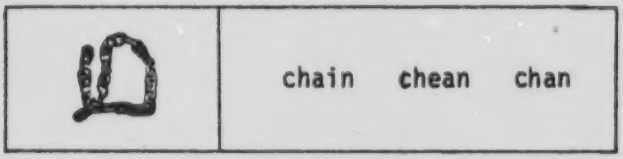
Revisions: The consonant blend distractor was dropped. Instead the third choice used a different vowel, sometimes with final e and sometimes without. The number of items was increased to ten. A new sample item is below:



D5. Decoding - Vowel digraphs

Tests the reader's ability to decode common vowel digraphs, including ee, ea, oa, and ai. The task is to select the word which names a picture. One incorrect choice has a single vowel; the other has an incorrect vowel combination.

Sample:



Pilot Data:

Number of Items	7
Number of Students	69
Mean Total Correct	5.67
Standard Deviation	1.86
Mean Proportion Correct	.81
Internal Consistency	.81

Ability-OTL Matrix:

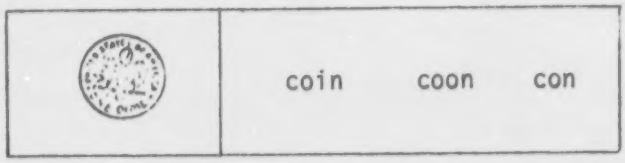
MEAN / COUNT	1	2	3	4	5	ROW TOTAL
VOCSTRAT						
1 LOW STRATA	3.0 / 9	0.0 / 0	4.0 / 1	6.8 / 6	5.5 / 4	4.7 / 20
2 MIDDLE STRATA	5.0 / 1	0.0 / 0	5.8 / 10	7.0 / 1	6.6 / 8	6.1 / 20
3 HIGH STRATA	0.0 / 0	7.0 / 8	7.0 / 1	6.8 / 8	0.0 / 0	6.9 / 17
COLUMN TOTAL	3.2 / 10	7.0 / 8	5.8 / 12	6.8 / 15	6.3 / 12	5.9 / 57

Revisions: The total number of items was expanded to 10. One old item was dropped. In two new items the correct response is the short vowel.

D6. Decoding - Vowel diphthongs

Tests the reader's ability to decode common vowel diphthongs, including ou, oi, aw, and oy. Also includes oo as in moon; ow as in cow. Two items with ey as in key.

Sample:



Pilot Data:

Number of Items	12
Number of Students	75
Mean Total Correct	9.59
Standard Deviation	2.55
Mean Proportion Correct	.80
Internal Consistency	.80

Ability-OTL Matrix:

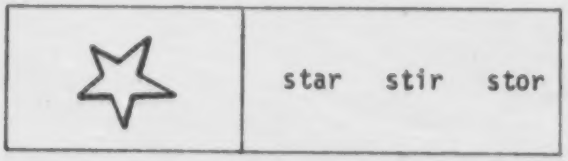
MEAN COUNT / VOCSTRAT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA	6.1 9	0.0 0	11.0 6	9.4 5	0.0 0	8.4 20
2 MIDDLE STRATA	8.0 1	0.0 0	9.0 2	10.4 14	10.7 3	10.1 20
3 HIGH STRATA	0.0 0	12.0 2	11.5 14	10.0 1	0.0 0	11.5 17
COLUMN TOTAL	6.3 10	12.0 2	11.1 22	10.1 20	10.7 3	9.5 57

Revisions: Distractors which involved a different spelling of the same sound were eliminated (au, ou), to make the test more closely one of letter-sound correspondence knowledge. The ey items were deleted. One item was replaced because of picture interpretation difficulties (straw). The replacement item is "book".

D7. Decoding - Vowel with r

Tests the reader's ability to decode a vowel when the sound of the vowel is influenced by the following consonant. Only words with an r following the vowel are included.

Sample:



Pilot Data:

Number of Items	12
Number of Students	75
Mean Total Correct	9.61
Standard Deviation	2.53
Mean Proportion Correct	.80
Internal Consistency	.79

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	4	5	ROW TOTAL
1 LOW STRATA	5.3 9	10.1 10	0.0 0	9.0 1	7.9 20
2 MIDDLE STRATA	10.0 2	9.9 9	0.0 0	10.3 9	10.1 20
3 HIGH STRATA	11.0 8	11.8 6	12.0 2	11.0 1	11.4 17
COLUMN TOTAL	8.2 19	10.4 25	12.0 2	10.3 11	9.7 57

Revisions: Two items omitted. One because of duplication of pattern. One because it showed a decline with increased OTL (unlike other items; may be a picture interpretation problem (marbles)).

D8. Decoding - Internal pattern in multisyllabic words

The reader must use his knowledge of letter-sound correspondences to decode more complex words. All words are of two syllables or more. Changes distinguishing one choice from another occur primarily in middle position. Reversal of letter patterns also occurs in distractors.

Sample:

palace palate please

Pilot Data:

Number of Items	10
Number of Students	75
Mean Total Correct	7.35
Standard Deviation	2.48
Mean Proportion Correct	.74
Internal Consistency	.78

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	5.5 13	7.0 7	0.0 0	0.0 0	6.0 20
2 MIDDLE STRATA	7.6 19	8.3 10	0.0 0	0.0 0	7.9 20
3 HIGH STRATA	8.6 8	9.0 5	10.0 2	10.0 2	9.1 17
COLUMN TOTAL	7.0 31	8.0 22	10.0 2	10.0 2	7.6 57

Revisions: None

D9. Consonant substitution with initial consonant blends

Tests knowledge of sounds associated with consonant blends. Also tests the student's ability to blend sounds. The student must select the initial blend or digraph which will combine with a VC ending to produce a real word. Some of the items are cross-referenced with subtest D2 and C1. If the student ignores the second half of the consonant blend, he will systematically choose the wrong response.

Sample:

hop __op	dr tw sk

Pilot Data:

Number of Items	10
Number of Students	75
Mean Total Score	7.11
Standard Deviation	2.78
Mean Proportion Correct	.71
Internal Consistency	.83

Ability-OTL Matrix:

VOCSTRAT / MEAN COUNT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA	0.0 0	7.0 10	9.0 0	5.0 1	3.8 9	5.6 20
2 MIDDLE STRATA	9.0 1	7.2 6	9.0 3	8.6 9	0.0 1	7.8 20
3 HIGH STRATA	9.0 8	9.3 8	0.0 0	10.0 1	0.0 0	9.2 17
COLUMN TOTAL	9.0 9	7.8 24	9.0 3	8.7 11	3.4 10	7.5 57

Revisions: None

D10. Decoding - Recognition of silent letters

Tests the reader's ability to decode a word when one letter is silent. The reader must identify the silent letter in a word. Common silent consonant patterns are included (mb, kn, and wr). Vowel combinations are also represented, including silent e in the final position in a word, and vowel digraphs (ea).

Sample:

comb	c o m b

Pilot Data:

Number of Items	12
Number of Students	75
Mean Total Correct	7.44
Standard Deviation	3.73
Mean Proportion Correct	.62
Internal Consistency	.88

Ability-OTL Matrix:

	MEAN COUNT	1	2	3	4	5	ROW TOTAL
VOCSTRAT							
1 LOW STRATA		4.1	6.5	3.0	0.0	0.0	5.3
		9	10	1	0	0	20
2 MIDDLE STRATA		7.0	8.3	9.1	0.0	2.7	7.8
		1	6	10	0	3	20
3 HIGH STRATA		0.0	11.0	10.7	12.0	0.0	11.1
		0	12	3	2	0	17
COLUMN TOTAL		4.4	8.8	9.0	12.0	2.7	7.9
		10	28	14	2	3	57

Revisions: This scale was dropped. Consultant feedback indicated that this is an area of confusion linguistically. This subtest was replaced with a Sight Words subtest from the BTES Phase II battery.

D11. Decoding - Matching variant consonant sounds.

Tests the reader's ability to decode both the sounds commonly associated with a single letter. The consonants included are c (cat vs. city) and g (goat vs. giant). The task requires the student to decode three words and circle the word in which the consonant sound matches the sound in a target word.

Sample:

cat

come center chop

Pilot Data:

Number of Items	7
Number of Students	61
Mean Total Score	4.46
Standard Deviation	1.89
Mean Proportion Correct	.64
Internal Consistency	.65

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA	0.0 0	3.5 4	5.0 6	7.0 1	3.4 9	4.1 20
2 MIDDLE STRATA	4.5 4	4.2 5	3.0 1	4.1 5	3.0 1	4.1 20
3 HIGH STRATA	4.5 8	0.0 0	6.8 8	5.0 1	0.0 0	5.6 17
COLUMN TOTAL	4.5 12	3.9 9	5.8 15	4.5 11	3.4 10	4.5 57

Revisions: The format was changed to that of the grade 5 variant consonants test. Instead of the auditory match to a word with picture, the test was restricted to sounds of c and symbols are introduced to represent the sounds. This eliminates the confusion of changing the task with a picture format.

Sample:

cup k s

C1. Context Clues - Initial consonant choice.

Tests the reader's ability to comprehend a sentence unit. A sentence is written with a word missing. Three choices are given to fill the blank, and the choices vary in initial consonant. The correct choice is always a consonant blend or digraph, and the types of distractors are the same as in the decoding test, D2. This provides a separate assessment of the student's ability to decode the response words. The additional skill tested in this scale is the reader's ability to process the sentence context and choose the word which completes the meaning of the sentence.

Sample:

> The _____ on the playground is broken.
>
> swing
>
> sing
>
> sting

Pilot Data:

Number of Items	10
Number of Students	75
Mean Total Score	7.53
Standard Deviation	3.12
Mean Proportion Correct	.75
Internal Consistency	.90

Ability-OTL Matrix:

MEAN / COUNT	1	2	5	ROW TOTAL
VOCSTRAT				
1 LOW STRATA	0.0 / 0	7.2 / 10	3.7 / 10	5.4 / 20
2 MIDDLE STRATA	8.9 / 4	9.0 / 6	8.4 / 10	8.6 / 20
3 HIGH STRATA	9.8 / 8	9.8 / 6	9.7 / 3	9.8 / 17
COLUMN TOTAL	9.4 / 12	8.4 / 22	6.5 / 23	7.9 / 57

Revisions: As in D2, some choices were nonsense words. This would make the task easier, since a nonsense word could not fit in context. Distractors were changed so that all were real words, even if the single consonant choice had to be eliminated.

C2. Word Structure - Syntax choices

As in C1, the reader must understand a sentence unit, and use his understanding to fill in a missing component. In this scale, the choices differ in that only one choice fits into the grammatical structure of the sentence. Other choices are common associates of the key content word(s), but they do not fit the syntax of the sentence. If the reader has understood the sentence, only one choice makes any sense.

Sample:

A _____ lives in the water.

swim

fish

under

Pilot Data:

Number of Items	5
Number of Students	69
Mean Total Score	4.23
Standard Deviation	1.43
Mean Proportion Correct	.85
Internal Consistency	.85

Ability-OTL Matrix:

MEAN COUNT / VCCSTRAT	1	3	4	ROW TOTAL
1 LOW STRATA	0.0 0	3.7 20	0.0 0	3.7 20
2 MIDDLE STRATA	4.8 4	4.8 16	0.0 0	4.8 20
3 HIGH STRATA	4.9 8	4.9 7	5.0 2	4.9 17
COLUMN TOTAL	4.8 12	4.3 43	5.0 2	4.4 57

Revision: This scale was expanded to 10 items.

C3. Context Clues - Word structure choices

As in C1 and C2, the reader must use context to select the best word to fill a blank. In this scale, the choices are all forms of the same base word. The reader must understand the meaning of common prefixes and suffixes, and know how to use them appropriately. By comparing performance on this scale to performance on C1 and C2, an estimate of the reader's understanding of affixes can be obtained.

Sample:

> Mary helps her mother at home.
> Mary is a very _____ girl.
>
> helper
>
> helpful
>
> helpless

Pilot Data:

Number of Items	10
Number of Students	69
Mean Total Score	5.80
Standard Deviation	2.68
Mean Proportion Correct	.58
Internal Consistency	.75

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	2.8 9	5.7 10	6.0 1	0.0 0	4.4 20
2 MIDDLE STRATA	1.0 1	6.1 7	5.7 12	0.0 0	5.6 20
3 HIGH STRATA	7.9 8	9.7 6	9.0 1	8.0 2	8.6 17
COLUMN TOTAL	4.9 18	6.9 23	5.9 14	8.0 2	6.1 57

Revisions: One item was substituted to reduce the attention to possessives and increase the weighting of comparative-superlative forms.

C4. - Context Clues - Pronouns

In the context clues format, the reader must choose the correct pronoun to complete a sentence. The choices are all personal pronouns, and vary in number and/or gender.

Sample:

> Susan ran to her house.
> Then _____ ran inside.
>
> she
>
> he
>
> it

Pilot Data:

Number of Items	7
Number of Students	75
Mean Total Score	4.79
Standard Deviation	1.90
Mean Proportion Correct	.68
Internal Consistency	.73

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	4.9 10	3.6 5	1.0 1	0.0 0	4.1 20
2 MIDDLE STRATA	5.0 6	4.0 5	5.0 9	0.0 0	4.8 20
3 HIGH STRATA	6.4 14	0.0 0	5.0 1	7.0 2	6.4 17
COLUMN TOTAL	5.6 30	3.7 14	4.6 11	7.0 2	5.0 57

Revisions: This scale was dropped. Scale M3, pronoun reference, was judged to be closer to the underlying reading skill and was expanded.

W1. Word Structure - Identification of root words

Tests the student's ability to identify the root word in a derived or inflected form. Either a prefix or a suffix may be added. Some words involve spelling changes, such as double consonants or changing y to i. In most items, at least one incorrect choice is another real word contained in the derived form.

Sample:

smartest	smart test art

Pilot Data:

Number of Items	16
Number of Students	69
Mean Total Score	10.17
Standard Deviation	4.55
Mean Proportion Correct	.64
Internal Consistency	.89

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	3	4	5	ROW TOTAL
1 LOW STRATA	4.4 9	9.0 1	0.0 0	12.2 10	8.5 20
2 MIDDLE STRATA	9.0 1	11.9 9	0.0 0	10.2 10	10.9 20
3 HIGH STRATA	0.0 0	10.0 1	15.5 2	13.1 14	12.2 17
COLUMN TOTAL	4.9 10	11.5 11	15.5 2	12.0 34	10.8 57

Revisions: A root word with prefix was added as an example. The word shelves was deleted due to a low total score and teacher reports of little coverage. It was replaced with an additional double-consonant word (cutting).

W2. Word Structure - Division of compound words

The student must identify the two parts of a compound word by drawing a line through the word to separate the parts.

Sample:

gold/fish

Pilot Data:

Number of Items	10
Number of Students	61
Mean Total Score	8.61
Standard Deviation	2.77
Mean Proportion Correct	.86
Internal Consistency	.94

Ability-OTL Matrix:

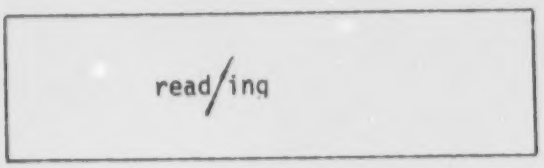
MEAN COLNT / VCCSTRAT	1	2	4	5	ROW TOTAL
1 LOW STRATA	6.1 9	13.8 2	14.7 6	0.0 0	10.6 20
2 MIDDLE STRATA	12.0 1	14.7 14	15.0 2	14.3 3	14.5 20
3 HIGH STRATA	0.0 0	14.1 9	14.9 8	0.0 0	14.5 17
COLUMN TOTAL	6.7 10	14.4 28	14.8 16	14.3 3	13.1 57

Revisions: None

W3. Word Structure - Division of root word from affix

Combines the content of W1, identification of root words, with the format of W2, drawing a line through a word. Only words without spelling changes are included. Assesses the reader's visual analysis of the word.

Sample:



Pilot Data:

Number of Items	12
Number of Students	61
Mean Total Score	7.84
Standard Deviation	3.54
Mean Proportion Correct	.65
Internal Consistency	.88

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	3	4	5	ROW TOTAL
1 LOW STRATA	4.6 9	14.0 1	0.0 0	17.5 10	11.5 20
2 MIDDLE STRATA	16.0 1	18.7 9	0.0 0	15.7 10	17.0 20
3 HIGH STRATA	0.0 0	10.0 1	25.5 2	22.1 14	21.8 17
COLUMN TOTAL	5.7 10	17.5 11	25.5 2	18.9 34	16.5 57

Revisions: A more familiar word with an re— prefix was substituted (replay for rewrite) because of the low score (22 percent correct). An example with a prefix was added.

W4. Word Structure - Number of syllables

The reader must identify the number of syllables in a word. Tests word reading as well visual-auditory segmentation. Identification of the number of syllables rather than the syllables themselves allows for differences across curricula in exact syllabification rules (slip-pers and slipp-ers both yield 2 syllables).

Sample:

window	1	2	3

Pilot Data:

Number of Items	16
Number of Students	75
Mean Total Score	10.01
Standard Deviation	4.17
Mean Proportion Correct	.63
Internal Consistency	.85

Ability-OTL Matrix:

MEAN / COUNT	1	3	4	5	ROW TOTAL
VCCSTRAT					
1 LOW STRATA	6.1 9	6.0 1	9.5 4	13.2 6	8.9 20
2 MIDDLE STRATA	7.0 1	8.8 5	11.0 5	9.2 5	9.3 20
3 HIGH STRATA	0.0 0	14.0 1	14.0 2	13.9 14	12.9 17
COLUMN TOTAL	6.2 10	9.0 11	11.0 11	12.8 25	10.6 57

Revisions: Four items were omitted to reduce scale size. Two items which were omitted had low scores, even for high ability student with high OTL. Two items duplicated patterns already contained in the scale.

W5. Word Structure - Contractions

The reader must indicate his understanding of contractions by matching a contraction with the correct expanded form and by selecting the appropriate contraction for a given phrase.

Samples:

He's mad at me.	I cannot go.
A. She	A. can't
B. He is	B. won't
C. It is	C. don't

Pilot Data:

Number of Items	4
Number of Students	76
Mean Total Score	2.67
Standard Deviation	1.49
Mean Proportion Correct	.67
Internal Consistency	.81

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	2	3	4	ROW TOTAL
1 LOW STRATA	1.8 10	2.3 9	1.0 7	1.8 26
2 MIDDLE STRATA	2.8 4	2.9 14	2.5 6	2.8 24
3 HIGH STRATA	4.0 5	3.5 13	4.0 2	3.6 22
COLUMN TOTAL	2.6 19	3.0 36	2.0 15	2.7 72

W5. Word Structure - Contractions Cont'd

Revisions: The context of a sentence was removed to make the test more closely reflect recognition of contractions. An additional distractor was added on all items. For all contractions except not (n't) the following choices are included:

1. Correct choice
2. Correct beginning, wrong ending
3. Wrong beginning, correct ending
4. Wrong beginning, wrong ending

Sample:

He's	She	He is
	It is	He was

A total of ten items were written in the new format.

M1. Word Meaning in Context - Multimeaning words

Tests the reader's ability to understand a sentence and to select the definition of a word which is appropriate to its use in that sentence. Tests the ability to deal appropriately with multiple meanings for a word.

Sample:

> The airplane rose in the air.
>
> A. a flower
>
> B. went up
>
> C. a pink color

Pilot Data:

Number of Items	10
Number of Students	76
Mean Total Score	6.55
Standard Deviation	2.60
Mean Proportion Correct	.66
Internal Consistency	.78

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	ROW TOTAL
1	4.3	0.0	4.3
LOW STRATA	26	0	26
2	6.8	8.1	7.2
MIDDLE STRATA	16	8	24
3	8.2	8.8	8.4
HIGH STRATA	13	9	22
COLUMN TOTAL	5.9	8.5	6.5
	55	17	72

Revisions: One item was taken as an example and replaced with a new item. One item with poor reactivity was replaced.

M2. Word Meaning in Context - Unfamiliar words

Tests the reader's ability to understand connected prose and to use context to define an unfamiliar word.

Sample:

> The old witch was cruel. She hit the little dog with a stick.
>
> A. nice
>
> B. mean
>
> C. friendly

Pilot Data:

Number of Items	5
Number of Students	76
Mean Total Score	3.65
Standard Deviation	-
Mean Proportion Correct	.73
Internal Consistency	-

Ability-OTL Matrix:

MEAN / COUNT	VOCSTRAT 1	VOCSTRAT 2	VOCSTRAT 3	ROW TOTAL
LOW STRATA	2.3 / 22	2.0 / 3	2.0 / 1	2.3 / 26
MIDDLE STRATA	3.9 / 13	4.8 / 8	4.7 / 3	4.3 / 24
HIGH STRATA	4.2 / 9	4.8 / 9	4.8 / 4	4.5 / 22
COLUMN TOTAL	3.2 / 44	3.3 / 20	4.4 / 8	3.6 / 72

Revisions: One item was revised to make the choices directly substitutable in the sentence ("to break" changed to "break").

M3. Word Meaning in Context - Pronoun reference

Pronouns are words which only take on specific meaning within a context. An important part of understanding connected discourse and of using context to determine meaning is the identification of pronoun referents.

Sample:

> Barbara pulled her wagon home.
> She had groceries for her mother.
>
> A. Mother
>
> B. Barbara
>
> C. the wagon

Pilot Data:

Number of items	5
Number of Students	76
Mean Total Score	2.67
Standard Deviation	1.48
Mean Proportion Correct	.53
Internal Consistency	.58

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	1.9 7	1.6 17	0.0 0	1.7 26
2 MIDDLE STRATA	2.5 4	3.1 16	2.8 4	2.9 24
3 HIGH STRATA	3.3 7	3.5 6	4.4 7	3.8 22
COLUMN TOTAL	2.6 18	2.5 43	3.8 11	2.7 72

Revisions: The scale was expanded to 10 items. Examples were added. Fewer items with quotations were included, in an effort to improve internal consistency. Some of the new items have an additional distractor. Items now range in difficulty, with easier items answered on the basis of gender alone, while harder items must be answered from context.

T. Text - Comprehension of prose (general discussion)

The prose comprehension subtests are composed according to a factorial design. Comprehension is defined in terms of two dimensions. One dimension is the type of information being processed or probed for in comprehension questions. The other dimension is relationship of the question to the text. This second dimension refers to the type of cognitive processing which the reader must go through to produce the desired response. These dimensions, and the levels of each dimension, are shown in the figure below:

	Relationship of Question to Text			
	Literal		Inferential	
Type of Information	Recall	Translation	Integration	Extrapolation
Event Detail				
Description				
Relationships				
Main Idea				

At each grade level, the comprehension test includes at least two items in each cell of the matrix. Eight passages are used at grade 2; seven passages are used at grade 5. Question types are not balanced across passages.

For the purposes of analyzing and reporting reactivity, this matrix is represented as four major scales. The scales are the four process categories. That is, scale T1 is Literal Recall, and is composed of items for each type of information. The matrix is reported in this way because of the belief that the processing dimension is particularly important. This dimension is also the least self-evident, and reporting these categories provides an opportunity for defining the categories and giving examples.

On the following page is a story used to provide examples in the scale description.

Oh, how Peter wished he could whistle! Peter saw his friend Sam playing with a dog. Whenever Sam whistled, the dog ran straight to him. Peter wished he could do that trick with his own dog. Willie. Peter tried and tried to whistle, but he just couldn't.

Peter went into his house and put on his father's old hat, to make himself feel more grown-up. He looked into the mirror to practice whistling. Still no whistle!

The next day Peter went outside to play. He sat on the front steps and tried to whistle. Then Peter saw his dog coming. Quick as a wink, Peter hid behind the stairs. He wanted to surprise Willie with a whistle. Peter puffed up his cheeks. He blew and blew and blew. Suddenly, out came a real whistle. Willie stopped and looked around to see who was making the noise.

"It's me," Peter shouted. He jumped out from behind the stairs. Willie raced straight up to him.

T1. Comprehension of Text - Literal, Recall

Questions of this type require the reader to remember information when the prompt is stated exactly as it was in the text. No rewording is required. The reader can use surface structure similarity as an aid in responding.

Samples:

> What did Peter wish he could do?
>
> (Answer: whistle; type - main idea)
>
> Where did Peter hide?
>
> (Answer: behind the stairs; type - description (setting))

Pilot Data:

Number of Items	9
Number of Students	76
Mean Total Score	5.46
Standard Deviation	2.35
Mean Proportion Correct	.61
Internal Consistency	.73

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	3	4	5	ROW TOTAL
1 LOW STRATA	3.8 4	3.3 22	0.0 0	3.4 26
2 MIDDLE STRATA	6.4 11	5.0 9	7.8 4	6.1 24
3 HIGH STRATA	7.6 13	5.7 7	8.5 2	7.1 22
COLUMN TOTAL	6.6 28	4.2 38	8.0 6	5.4 72

Revisions: None

T2. Comprehension of Text - Literal, translation

In these questions the reader must deal with rewording of information from the passage. The information is directly present in the text, and in a single sentence, but the reader must recognize the idea stated in different words. He must paraphrase the question or the responses to recognize the connection.

Samples:

> What happened after Sam whistled?
>
> (Answer: a dog ran up to Sam; type - relationship)
>
> What did Peter wear?
>
> (Answer: a hat of his father's; type - event detail)

Pilot Data:

Number of Items	9
Number of Students	76
Mean Total Score	4.99
Standard Deviation	2.78
Mean Proportion Correct	.55
Internal Consistency	.77

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	4	5	ROW TOTAL
VOCSTRAT						
1 LOW STRATA	2.9 14	3.0 6	3.0 3	5.0 1	0.0 0	3.0 26
2 MIDDLE STRATA	4.8 6	6.0 8	5.3 3	3.7 3	6.8 4	5.5 24
3 HIGH STRATA	4.3 7	7.3 4	0.0 0	7.5 4	7.7 7	6.5 22
COLUMN TOTAL	3.7 27	5.1 20	4.2 6	5.8 8	7.4 11	4.9 72

Revisions: None

T3. Comprehension of Text - Inferential, Integration

In these questions the reader must integrate information from different locations in a text. He must synthesize ideas. Inferences directly implied within a text are included in this category. Also included are questions where the information is in two contiguous sentences and the reader must understand the role of sentence order in implying connections (see example 1 below). In other questions pronouns or general words refer to elements elsewhere in the text and the reader must make the appropriate referential connections in order to integrate the material (see example 2 below).

Samples:

> Why did Peter hide behind the stairs?
>
> (Answer: so he could surprise Willie; type - relationship)
>
> What trick did Peter want to do with his dog?
>
> (Answer: whistle to call Willie; type - event detail)

Pilot Data:

Number of Items	10
Number of Students	76
Mean Total Score	5.03
Standard Deviation	2.54
Mean Proportion Correct	.50
Internal Consistency	.70

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	3.4 14	3.3 8	3.5 4	0.0 0	3.3 26
2 MIDDLE STRATA	3.9 10	5.3 4	5.3 6	6.8 4	5.0 24
3 HIGH STRATA	6.0 9	5.3 4	7.3 4	8.6 7	7.0 22
COLUMN TOTAL	4.2 33	4.1 16	5.4 14	7.9 11	5.0 72

Revisions: None

T4. Comprehension of Text - Inferential, Extrapolation

These questions require the reader to relate the text to his own experience and knowledge. He must supply from experience something not present in the text. The most frequent kind of question is one which requires the reader to predict what might happen next, to go beyond the information given. (Questions of this type are the most open to alternative interpretations. Care was taken to try to write questions which would relate to common experiences.)

Samples:

> How do you think Peter felt at the end of the story?
>
> (Answer: happy; type - description (character))
>
> What lesson can we learn from this story?
>
> (Answer: keep trying and you will learn to do something (If at first you don't succeed.....); type - main idea)

Pilot Data:

Number of Items	10
Number of Students	76
Mean Total Score	5.37
Standard Deviation	2.46
Mean Proportion Correct	.54
Internal Consistency	.68

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
VOCSTRAT 1	3.2	0.0	3.3	3.2
LOW STRATA	22	0	4	26
2	5.5	7.5	5.3	5.8
MIDDLE STRATA	10	4	10	24
3	6.2	8.5	7.8	7.2
HIGH STRATA	9	2	11	22
COLUMN TOTAL	4.4	7.0	6.1	5.3
	41	6	25	72

Revisions: None

III A. Comprehension - Applied

This scale tests the reader's ability to deal with material other than straight prose. Many of the same cognitive skills of comprehension are probably involved, but the reader must apply his skills to a specific task or format. Many of the items require the reader to search through or scan the page to locate information. Other items test the reader's orientation to the material, his understanding of what kind of information is being presented. The following components are included in the scale (with a separate OTL matrix for each):

1. Letters (1 letter - 3 questions)
2. Signs (selection of appropriate sign - 4 items)
3. Directions (understanding directions - 2 directions, 5 questions; following directions - 6 items)
4. Picture interpretation (1 picture - 5 questions)
5. Table of contents (1 table - 3 questions)
6. Riddles (3)
7. Schedule (1 schedule - 2 questions)

Pilot Data (total scale):

Number of Items	31
Number of Students	76
Mean Total Correct	20.07
Standard Deviation	6.56
Mean Proportion Correct	.65
Internal Consistency	.89

1. LETTERS

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
VOCSTRAT 1	0.9	0.7	0.0	0.8
LOW STRATA	17	9	0	26
2	1.8	1.3	2.5	1.8
MIDDLE STRATA	14	6	4	24
3	1.7	3.0	2.7	2.3
HIGH STRATA	11	4	7	22
COLUMN TOTAL	1.4	1.4	2.6	1.6
	42	19	11	72

Revisions: None

2. SIGNS

Ability-OTL Matrix:

VOCSTRAT	MEAN COUNT	1	2	ROW TOTAL
1 LOW STRATA		2.3 7	1.9 19	2.0 26
2 MIDDLE STRATA		3.6 8	3.1 16	3.3 24
3 HIGH STRATA		3.2 9	3.5 13	3.4 22
COLUMN TOTAL		3.1 24	2.7 48	2.8 72

Revisions: One item was deleted due to high score with OTL=1, and subsequent lower scores (principal's door). One item was simplified to reduce ability differences due to unfamiliar content (ham radio, Heathkit). One item had two distractors replaced to reduce the role of immediate association and of unfamiliar vocabulary.

3. DIRECTIONS

Ability-OTL Matrix:

VOCSTRAT	MEAN COUNT	1	2	3	4	ROW TOTAL
1 LOW STRATA		5.1 17	5.3 6	5.0 3	0.0 0	5.2 26
2 MIDDLE STRATA		6.7 14	6.3 3	7.3 3	9.0 4	7.1 24
3 HIGH STRATA		7.0 11	8.8 4	0.0 0	9.9 7	8.6 22
COLUMN TOTAL		6.4 42	6.0 13	6.2 6	9.5 11	6.9 72

Revisions: Minor editing to remove sex bias.

4. PICTURE INTERPRETATION

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	2	3	ROW TOTAL
1 LOW STRATA	2.3 19	2.7 7	2.4 26
2 MIDDLE STRATA	4.3 16	3.6 8	4.0 24
3 HIGH STRATA	4.4 13	4.3 9	4.4 22
COLUMN TOTAL	3.5 48	3.6 24	3.5 72

Revisions: Editing for clarity. One item changed to depend on picture interpretation more than background knowledge.

5. TABLE OF CONTENTS

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	1.4 7	3.0 1	1.3 18	1.4 26
2 MIDDLE STRATA	2.5 8	1.3 3	2.0 13	2.1 24
3 HIGH STRATA	2.9 9	3.0 4	3.0 9	3.0 22
COLUMN TOTAL	2.3 24	2.4 8	1.9 40	2.1 72

Revisions: Edited for clarification.

6. RIDDLES

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	5	ROW TOTAL
1 LOW STRATA	1.8 26	1.8 26
2 MIDDLE STRATA	2.7 24	2.7 24
3 HIGH STRATA	2.8 22	2.8 22
COLUMN TOTAL	2.4 72	2.4 72

Revisions: Expanded from 3 to 5 items.

7. SCHEDULE

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	1	2	ROW TOTAL
1 LOW STRATA	0.4 18	0.1 8	0.3 26
2 MIDDLE STRATA	0.5 17	1.0 7	0.6 24
3 HIGH STRATA	0.7 15	0.9 7	0.8 22
COLUMN TOTAL	0.5 50	0.5 22	0.6 72

Revision: Dropped from scale due to difficulty and low OTL.

Grade 5 Reading

D1. Decoding - Variant Consonants

Test the reader's ability to decode words containing a variant consonant. Symbols are introduced to represent the two possible sounds associated with a letter (c → k, s; g → g, j). The reader must identify the symbol which represents the sound associated with a specified consonant in a word.

Sample:

caught	k	s	g	j

Pilot Data:

Number of Items	18
Number of Students	80
Mean Total Score	14.61
Standard Deviation	3.11
Mean Proportion Correct	.81
Internal Consistency	.79

Ability-OTL Matrix (Language):

VOCSTRAT / MEAN COUNT	0	3	4	ROW TOTAL
1 LOW STRATA	11.800	13.375	12.100	11.435
	5	8	10	23
2 MIDDLE STRATA	15.455	15.182	14.250	15.154
	11	11	4	26
3 HIGH STRATA	16.400	16.000	16.333	16.250
	10	8	6	24
COLUMN TOTAL	15.115	14.000	13.800	14.342
	26	27	20	73

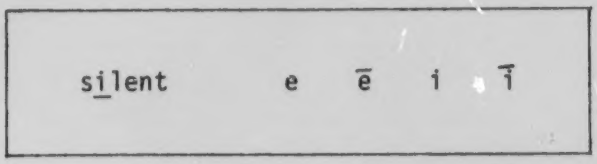
Revisions: The scale was reduced from 18 items to 12. Items varied in reactivity to instruction, and items with a positive slope were retained. (The word "guest", for example, showed 38% correct at OTL=3 and 90% correct at OTL=4 for low ability students.) One new word was added to give an equal number of words with each sound.

* Note: An OTL rating of 0 means that no rating was available.

D2. Decoding - Long-Short Vowel Sounds

Tests the reader's ability to decode vowels in words which follow common patterns. Patterns tested include (1) CVC short vowel pattern, (2) CVCe long vowel pattern, (3) CV long vowel in accented syllable. Both real words and nonsense words are included for all three patterns. Sounds of y and irregular patterns are also tested, using real words only. As in D1, symbols are introduced to represent the sounds (short e = e, long e = $\bar{e}$, short i = i, long i = $\bar{i}$). The reader must identify the symbol which represents the sound associated with a specified vowel in a word.

Sample:



Pilot Data:

Number of Items	32
Number os Students	80
Mean Total	21.94
Standard Deviation	5.66
Mean Proportion Correct	.69
Internal Consistency	.83

Ability-OTL Matrix (Language):

MEAN COUNT / VOCSTRAT	0	3	4	ROW TOTAL
1 LOW STRATA	15.200 5	14.875 8	17.000 10	15.870 23
2 MIDDLE STRATA	24.091 11	20.273 11	25.500 4	22.692 26
3 HIGH STRATA	26.000 10	26.000 8	23.667 6	25.417 24
COLUMN TOTAL	23.115 26	20.370 27	20.700 20	21.438 73

Revisions: A number of items were dropped from the scale. Responses from consultants and student error rates suggested that a number of items were ambiguous, particularly for the long vowel in accented syllable rule. Items with low reactivity were also dropped. Two new words were added to round out the testing of different rules and sounds.

D3. Decoding - Vowel with r

Tests the reader's ability to decode words where the vowel is followed by r. As in D1 and D2, symbols are introduced to represent sounds (fair - air, near - ear, her - er, car - ar). The reader must identify which of the four sounds occurs in a word. Words include regular and irregular patterns.

Sample:

stare	air	ear	er	ar

Pilot Data:

Number of Items	22
Number of Students	80
Mean Total Score	15.28
Standard Deviation	4.95
Mean Proportion Correct	.69
Internal Consistency	.87

Ability-OTL Matrix (Language):

MEAN COUNT	0	2	5	ROW TOTAL
VOCSTRAT 1 LOW STRATA	10.800 5	11.600 10	9.250 8	10.609 23
2 MIDDLE STRATA	14.000 11	14.250 4	17.545 11	15.538 26
3 HIGH STRATA	18.800 10	19.500 6	17.875 8	18.667 24
COLUMN TOTAL	15.231 26	14.500 20	15.185 27	15.014 73

Revisions: It appeared from the reactivity data that items with the most positive pattern were those where the letters "ar" are associated with the sound air (as in careful). Most of these items were retained. Many other (unpromising) items were deleted to reduce the scale to 10 items.

D4. Decoding - Recognition of silent letters

Tests the reader's ability to decode a word when one letter is silent. The reader must identify the silent letter in a word. Common silent consonant patterns are included (kn, wr, gn, tch). Vowel combinations are also represented, including silent e and vowel digraphs.

Sample:

design	d e s i g n

Pilot Data:

Number of Items	16
Number of Students	80
Mean Total Score	12.07
Standard Deviation	4.77
Mean Proportion Correct	.75
Internal Consistency	.93

Ability-OTL Matrix (Language):

MEAN / COUNT	0	2	3	ROW TOTAL
VOCSTRAT				
1 LOW STRATA	6.000 5	9.900 10	6.875 8	8.000 23
2 MIDDLE STRATA	13.182 11	13.000 4	13.182 11	13.154 26
3 HIGH STRATA	14.000 10	14.167 6	14.500 8	14.208 24
COLUMN TOTAL	12.115 26	11.800 20	11.704 27	11.877 73

Revisions: This scale was dropped due to consultant advise about the ambiguity and imprecision of "silent letters," particularly from the linguistic perspective. Reactivity data were also unencouraging.

C1. Context Clues - Semantic Appropriateness

Tests the reader's ability to understand a sentence or paragraph and use the semantic context to fill in a missing element. A sentence or paragraph is written with a key word or words deleted. The reader must select the best word to fill each blank and complete the meaning of the passage. In the pilot version, 5 items use sentence context and 4 use paragraph context. This scale was adopted from the grade 5 reading ETU.

Sample:

> Because the ocean roars, we need to
>
> _____ to be heard.
>
> talk
>
> scream
>
> whisper

Pilot Data:

Number of Items	9
Number of Students	80
Mean Total Score	7.43
Standard Deviation	1.77
Mean Proportion Correct	.82
Internal Consistency	.69

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	2	3	4	ROW TOTAL
1 LOW STRATA	5.882 17	5.600 5	0.0 0	6.000 1	5.826 23
2 MIDDLE STRATA	7.857 7	7.875 16	9.000 1	9.000 2	8.000 26
3 HIGH STRATA	7.667 3	8.250 15	8.500 6	0.0 0	8.208 24
COLUMN TOTAL	6.593 27	7.694 36	8.571 7	8.000 3	7.384 73

Revisions: One more paragraph item added (with three blanks.)

C2. Context Clues - Word Structure

This scale uses the same format as C1. A sentence is written with a word deleted. The reader must process the sentence context to select the missing word. The response choices in this scale are all forms of the same base word. The reader must use his knowledge of word structure and of the meanings of prefixes and suffixes in order to select the appropriate word.

Sample:

> The _____ little dog fought the big dog who had stolen his bone.
>
> A. fearsome
> B. fearful
> C. fearless
> D. fearfully

Pilot Data:

Number of Items	10
Number of Students	80
Mean Total Score	6.79
Standard Deviation	2.40
Mean Proportion Correct	.68
Internal Consistency	.77

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	1	2	3	4	ROW TOTAL
1 LOW STRATA	4.529 17	0.0 0	4.000 1	3.500 4	8.000 1	4.478 23
2 MIDDLE STRATA	7.714 7	10.000 1	7.500 6	8.300 10	8.000 2	7.231 26
3 HIGH STRATA	7.667 3	8.167 6	8.111 9	8.000 6	0.0 0	8.042 24
COLUMN TOTAL	5.704 27	8.429 7	7.625 16	6.250 20	8.000 3	6.630 73

Revisions: One difficult and one easy item item were dropped. Since the suffix items showed greatest reactivity, two new suffix items were added.

C3. Context Clues - Pronouns

Tests the reader's understanding of pronoun usage. As in C1 and C2, a sentence is written with a word deleted. The reader must select the one of four pronouns which fits in the context. The scale includes both personal pronouns (he, they, her, them) and relative pronouns (who, what).

Sample:

> If you don't want those leftovers,
>
> I'll feed _____ to the cat.
>
> A. it
>
> B. they
>
> C. them
>
> D. you

Pilot Data:

Number of Items	10
Number of Student	80
Mean Total Score	7.70
Standard Deviation	2.39
Mean Proportion Correct	.77
Internal Consistency	.78

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	1	2	4	ROW TOTAL
1 LOW STRATA	5.824 17	0.0 0	3.200 5	10.000 1	5.435 23
2 MIDDLE STRATA	7.571 7	8.000 2	8.000 14	8.667 3	7.962 26
3 HIGH STRATA	9.333 3	9.333 3	8.750 12	9.333 6	9.042 24
COLUMN TOTAL	6.667 27	8.000 5	7.516 31	9.200 10	7.521 73

Revisions: None

W1. Word Structure - Identification of root words

Tests the student's ability to identify the root word in a derived or inflected form. Either a prefix, a suffix, or 2 affixes may be added. In most items, at least one incorrect choice is another real word contained in the derived form. In some items the reader must distinguish between a form with one affix removed (fattening - fatten) and the smallest root word (fattening - fat). Only the latter is counted correct.

Sample:

fattening	fat	ten	fatten

Pilot Data:

Number of Items	14
Number of Students	80
Mean Total Score	11.10
Standard Deviation	2.55
Mean Proportion Correct	.79
Internal Consistency	.78

Ability-OTL Matrix:

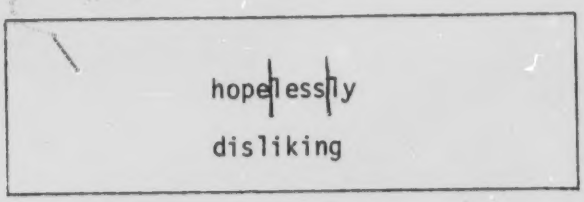
MEAN COUNT / VCCSTRAT	0	2	3	5	ROW TOTAL
1 LOW STRATA	9.235 17	10.000 1	6.250 4	8.000 1	8.696 23
2 MIDDLE STRATA	10.143 7	11.000 3	11.818 11	12.600 5	11.423 26
3 HIGH STRATA	12.667 3	12.571 7	12.417 12	13.000 2	12.542 24
COLUMN TOTAL	9.852 27	11.909 11	11.259 27	12.125 8	10.932 73

Revisions: The opinion of the consultants was that items like the sample, which require reduction to the smallest root, were unfair. The reader could understand the basic principle and still respond "incorrectly." The high rate of choosing partially-reduced distractors confirmed this criticism. Reactivity was low (slopes flat) for such items as well. So partially-reduced distractors (like "fatten") were eliminated and replaced with other words within the derived form.

W2. Word Structure - Division of root words and affixes

The reader must indicate word structure by drawing two lines through a word to separate the root word from a prefix and a suffix or to separate a root word and two suffixes. In comparison to W1, this format emphasizes the reader's visual analysis and word segmentation.

Sample:



Pilot Data:

Number of Items	13
Number of Students	80
Mean Total Score	8.99
Standard Deviation	3.40
Mean Proportion Correct	.69
Internal Consistency	.84

Ability-OTL Matrix:

MEAN / COUNT	0	2	3	4	5	ROW TOTAL
VOCSTRAT						
1 LOW STRATA	6.647 17	4.000 1	0.0 0	4.750 4	1.000 1	5.957 23
2 MIDDLE STRATA	8.429 7	7.667 3	9.727 11	11.333 3	10.500 2	9.385 26
3 HIGH STRATA	11.333 3	10.000 7	10.800 10	10.000 4	0.0 0	10.500 24
COLUMN TOTAL	7.630 27	8.818 11	10.238 21	8.455 11	7.333 3	8.671 73

Revisions: One item was deleted.

W3. Word Structure - First Syllable

The reader must identify the first syllable in a word of two or three syllables. Three choices are given.

Sample:

procession	A. pro	B. proc	C. proces

Pilot Data:

Number of Items	14
Number of Students	80
Mean Total Score	9.69
Standard Deviation	3.32
Mean Proportion Correct	.69
Internal Consistency	.82

Ability-OTL Matrix:

MEAN / COUNT / VOCSTRAT	0	1	2	3	5	ROW TOTAL
1 LOW STRATA	6.471 17	0.0 0	8.000 1	5.750 4	12.000 1	6.652 23
2 MIDDLE STRATA	9.429 7	12.667 3	11.667 3	9.455 11	10.500 2	10.154 26
3 HIGH STRATA	12.667 3	12.500 2	12.857 7	10.833 12	0.0 0	11.792 24
COLUMN TOTAL	7.926 27	12.600 5	12.091 11	9.519 27	11.000 3	9.589 73

Revisions: None

W4. Word Structure - Last Syllable

The reader must identify the last syllable in a word of two to four syllables. Three choices are given.

Sample:

filament	A. ament	B. ment	C. ent

Pilot Data:

Number of Items	12
Number of Students	80
Mean Total Score	8.13
Standard Deviation	2.68
Mean Proportion Correct	.68
Internal Consistency	.73

Ability-OTL Matrix:

MEAN / COUNT	0	1	2	ROW TOTAL
VOCSTRAT				
1 LOW STRATA	6.471 17	9.000 1	7.000 5	6.696 23
2 MIDDLE STRATA	7.429 7	8.000 5	7.857 14	7.769 26
3 HIGH STRATA	8.667 3	6.000 4	9.684 19	9.417 24
COLUMN TOTAL	6.963 27	8.125 8	8.658 38	7.973 73

W4. Word Structure - Last Syllable Cont'd

Ability-OTL Matrix (Language):

MEAN CCUNT / VOCSTRAT	0	1	3	ROW TOTAL
1 LOW STRATA	7.000 5	6.300 10	7.000 8	6.696 23
2 MIDDLE STRATA	7.455 11	9.000 4	7.636 11	7.769 26
3 HIGH STRATA	10.200 10	8.000 6	9.500 8	9.417 24
CCLUMN TOTAL	8.423 26	7.350 20	8.000 27	7.973 73

Revisions: One item was deleted because of a detrimental confusion between the last syllable ble and the suffix able.

W5. Word Structure - Number of Syllables

The reader must identify the number of syllables in a word. Tests word reading as well as visual-auditory segmentation. Identification of the number of syllables rather than the syllables themselves allows for differences in exact syllabification rules, an area of some uncertainty linguistically.

Sample:

exactly	1	2	3	4	5

Pilot Data:

Number of Items	15
Number of Students	80
Mean Total Score	11.49
Standard Deviation	3.09
Mean Proportion Correct	.77
Internal Consistency	.82

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	1	2	3	5	ROW TOTAL
1 LOW STRATA	9.353 17	0.0 0	14.000 1	6.750 4	12.000 1	9.217 23
2 MIDDLE STRATA	14.714 7	12.333 3	12.667 3	12.000 11	13.000 2	12.115 26
3 HIGH STRATA	14.333 3	11.500 2	12.571 7	12.083 12	0.0 0	12.458 24
COLUMN TOTAL	10.519 27	12.000 5	12.727 11	11.259 27	12.667 3	11.315 73

Revisions: Two ambiguous items were deleted (beautiful and miniature).

V2. Vocabulary - Definition of affixes or words with affixes.

This subtest represents the meaning component of the section on word structure. Prefixes and suffixes are useful elements to recognize in word identification because they are frequently occurring meaning units. An affix carries meaning with it and transforms the meaning of the base word in a predictable way. The meaning of many derived forms can be unlocked by knowing the meaning of the appropriate affix(es). This subtest assesses the reader's knowledge of the meaning of certain common affixes and his ability to define words with affixes. Both prefixes and suffixes are included.

Samples:

disbelieve	resell
A. completely	A. making a sale
B. take away	B. one who sells
C. partly	C. something sold
D. not	D. sell again

Pilot Data:

Number of Items	10
Number of Students	73
Mean Total Score	6.00
Standard Deviation	2.14
Mean Proportion Correct	.60
Internal Consistency	.67

Ability-OTL Matrix (affixes):

MEAN COUNT / VOCSTRAT	0	1	2	3	5	ROW TOTAL
1 LOW STRATA	1.909 11	0.0 0	2.000 1	2.750 4	2.000 5	2.095 21
2 MIDDLE STRATA	3.000 7	5.000 1	4.143 7	4.000 3	2.833 6	3.500 24
3 HIGH STRATA	3.400 5	4.250 4	4.500 6	3.500 2	3.333 3	3.900 20
COLUMN TOTAL	2.565 23	4.400 5	4.143 14	3.333 9	2.643 14	3.169 65

V2. (continued)

Ability-OTL Matrix (words with affixes):

VOCSTRAT	MEAN COUNT	0	1	2	3	4	5	ROW TOTAL
1 LOW STRATA		1.364	0.0	2.000	2.250	0.0	2.000	1.714
		11	0	1	4	0	5	21
2 MIDDLE STRATA		3.000	3.500	3.500	2.000	3.000	2.500	3.000
		7	2	6	3	4	2	24
3 HIGH STRATA		2.600	4.667	3.857	4.000	3.667	0.0	3.650
		5	3	7	2	3	0	20
COLUMN TOTAL		2.130	4.200	3.571	2.556	3.286	2.143	2.785
		23	5	14	9	7	7	65

Revisions: None

M1. Word Meaning in Context - Multimeaning words

Tests the reader's ability to understand a sentence and to select the definition of a word which is appropriate to its use in that sentence. Tests the ability to deal appropriately with multiple meanings for a word.

Sample:

> Roses cannot stand the cold.
>
> A. rise to one's feet
>
> B. stay in place
>
> C. a place for a small business
>
> D. bear, endure

Pilot Data:

Number of Items	10
Number of Students	73
Mean Total Score	5.34
Standard Devation	3.19
Mean Proportion Correct	.53
Internal Consistency	.83

Ability-OTL Matrix:

MEAN / COUNT	0	1	3	4	ROW TOTAL
VOCSTRAT					
1 LOW STRATA	2.273 / 11	4.600 / 5	2.000 / 1	4.000 / 4	3.143 / 21
2 MIDDLE STRATA	4.286 / 7	2.500 / 2	5.417 / 12	3.667 / 3	4.625 / 24
3 HIGH STRATA	7.600 / 5	0.0 / 0	8.077 / 13	8.500 / 2	8.050 / 20
COLUMN TOTAL	4.087 / 23	4.000 / 7	6.615 / 26	4.889 / 9	5.200 / 65

Revisions: Definitions were simplified to reduce the verbal processing and vocabulary load.

M2. Word Meaning in context - Unfamiliar words

As in M1, the reader must use the context of a sentence in order to understand the meaning of a word in that sentence. In this scale, one word in each sentence is an unfamiliar word (sometimes a nonsense word). The reader must process sentence context in order to select the best definition for the unknown word.

Sample:

> The car was so badly entrenched in the mud that we called a tow truck.
>
> A. stuck
> B. built
> C. dirty
> D. skidded

Pilot Data:

Number of Items	10
Number of Students	73
Mean Total Score	7.89
Standard Deviation	2.12
Mean Proportion Correct	.79
Internal Consistency	.72

Ability-OTL Matrix:

MEAN COUNT / VCCSTRAT	0	1	2	3	ROW TOTAL
1 LOW STRATA	5.636 11	8.000 5	7.000 5	0.0 0	6.524 21
2 MIDDLE STRATA	7.714 7	7.750 4	7.125 8	7.400 5	7.458 24
3 HIGH STRATA	8.400 5	10.000 3	9.400 5	9.143 7	9.150 20
COLUMN TOTAL	6.870 23	8.417 12	7.722 18	8.417 12	7.677 65

Revisions: None

M3. Word Meaning in context - pronoun reference

Pronouns are words which only take on specific meaning within a context. An important part of understanding connected discourse and of using context to determine meaning is the identification of pronoun referents. In this subtest two paragraphs are presented. In each paragraph several pronouns are underlined. The reader must indicate his understanding of pronoun reference by selecting the appropriate referent for each underlined pronoun.

Sample:

> Mother and I looked at my clothes. She said, "You need some new shirts, Pearl. We will buy them on Saturday."
>
> them means
>
> A. clothes
> B. shirts
> C. Mother
> D. Mother and Pearl

Pilot Data:

Number of Items	9
Number of Students	73
Mean Total Score	6.03
Standard Deviation	2.07
Mean Proportion Correct	.67
Internal Consistency	.66

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	1	2	3	4	ROW TOTAL
1 LOW STRATA	4.545 11	0.0 0	5.000 1	4.000 4	5.800 5	4.762 21
2 MIDDLE STRATA	6.000 7	8.000 1	5.556 9	4.333 3	5.750 4	5.667 24
3 HIGH STRATA	6.400 5	8.250 4	8.000 6	8.500 2	8.333 3	7.750 20
COLUMN TOTAL	5.391 23	8.200 5	5.438 16	5.111 9	6.417 12	6.015 65

Revisions: None

M4. Word Meaning in context - Figurative language

By grade 5, students are encountering writing containing figures of speech. Words and phrases are used with a non-literal meaning. Some such phrases are used in idiomatic expressions ("I saw the light."). Others are less familiar similes (see the example below). This subtest looks at the reader's ability to understand the meaning of such phrases in context. A literal translation is often offered as one distractor.

Sample:

> The soldier fought like a tiger to protect his home.
>
> A. in a striped uniform
> B. weakly and unwillingly
> C. with sharp claws
> D. bravely and fiercely

Pilot Data:

Number of Items	7
Number of Students	73
Mean Total Score	4.59
Standard Deviation	2.01
Mean Proportion Correct	.66
Internal Consistency	.74

Ability-OTL Matrix:

MEAN / COUNT	0	1	2	4	ROW TOTAL
VOCSTRAT					
1 LOW STRATA	2.000 11	0.0 0	2.800 5	4.400 5	2.762 21
2 MIDDLE STRATA	4.714 7	7.000 1	5.143 14	4.000 2	5.000 24
3 HIGH STRATA	5.000 5	5.500 4	5.636 11	0.0 0	5.450 20
COLUMN TOTAL	3.478 23	5.000 5	4.933 30	4.286 7	4.415 65

Revisions: None

T. Text - Comprehension of prose (general discussion)

The prose comprehension subtests are composed according to a factorial design. Comprehension is defined in terms of two dimensions. One dimension is the type of information being processed or probed for in comprehension questions. The other dimension is relationship of the question to the text. This second dimension refers to the type of cognitive processing which the reader must go through to produce the desired response. These dimensions, and the levels of each dimension, are shown in the figure below:

	Relationship of Question to Text			
	Literal		Inferential	
Type of Information	Recall	Translation	Integration	Extrapolation
Event Detail				
Description				
Relationships				
Main Idea				

At each grade level, the comprehension test includes at least two items in each cell of the matrix. Eight passages are used at grade 2; seven passages are used at grade 5. Question types are not balanced across passages.

For the purposes of analyzing and reporting reactivity, this matrix is represented as four major scales. The scales are the four process categories. That is, scale T1 is Literal Recall, and is composed of items for each type of information. The matrix is reported in this way because of the belief that the processing dimension is particularly important. This dimension is also the least self-evident, and reporting these categories provides an opportunity for defining the categories and giving examples.

On the following page is a story used to provide examples in the scale description.

Oh, how Peter wished he could whistle! Peter saw his friend Sam playing with a dog. Whenever Sam whistled, the dog ran straight to him. Peter wished he could do that trick with his own dog. Willie. Peter tried and tried to whistle, but he just couldn't.

Peter went into his house and put on his father's old hat, to make himself feel more grown-up. He looked into the mirror to practice whistling. Still no whistle!

The next day Peter went outside to play. He sat on the front steps and tried to whistle. Then Peter saw his dog coming. Quick as a wink, Peter hid behind the stairs. He wanted to surprise Willie with a whistle. Peter puffed up his cheeks. He blew and blew and blew. Suddenly, out came a real whistle. Willie stopped and looked around to see who was making the noise.

"It's me," Peter shouted. He jumped out from behind the stairs. Willie raced straight up to him.

T1. Comprehension of Text - Literal, Recall

Questions of this type require the reader to remember information when the prompt is stated exactly as it was in the text. No rewording is required. The reader can use surface structure similarity as an aid in responding.

Samples:

> What did Peter wish he could do?
>
> (Answer: whistle; type - main idea)
>
> Where did Peter hide?
>
> (Answer: behind the stairs; type - description (setting))

Pilot Data:

Number of Items	8
Number of Students	73
Mean Total Score	5.21
Standard Deviation	1.62
Mean Proportion Correct	.65
Internal Consistency	.47

Ability-OTL Matrix:

MEAN COUNT / VCCSTRAT	0	2	3	4	ROW TOTAL
1 LOW STRATA	3.909 11	0.0 0	4.200 5	5.000 5	4.238 21
2 MIDDLE STRATA	4.286 7	6.000 1	5.083 12	4.500 4	4.792 24
3 HIGH STRATA	7.000 5	6.500 4	6.250 8	6.667 3	6.550 20
COLUMN TOTAL	4.696 23	6.400 5	5.280 25	5.250 12	5.154 65

Revisions: None

T2. Comprehension of Text - Literal, translation

In these questions the reader must deal with rewording of information from the passage. The information is directly present in the text, and in a single sentence, but the reader must recognize the idea stated in different words. He must paraphrase the question or the responses to recognize the connection.

Samples:

> What happened after Sam whistled?
>
> (Answer: a dog ran up to Sam;
> type - relationship)
>
> What did Peter wear?
>
> (Answer: a hat of his father's;
> type - event detail)

Pilot Data:

Number of Items	8
Number of Students	73
Mean Total Score	4.66
Standard Deviation	1.92
Mean Proportion Correct	.58
Internal Consistency	.56

Ability-OTL Matrix:

MEAN / COUNT	0	2	4	ROW TOTAL
VOCSTRAT				
1 LOW STRATA	2.727 / 11	3.200 / 5	4.400 / 5	3.238 / 21
2 MIDDLE STRATA	4.143 / 7	4.667 / 9	5.125 / 8	4.667 / 24
3 HIGH STRATA	6.000 / 5	5.444 / 9	6.500 / 6	5.900 / 20
COLUMN TOTAL	3.870 / 23	4.652 / 23	5.368 / 19	4.585 / 65

Revisions: None

T3. Comprehension of Text - Inferential, Integration

In these questions the reader must integrate information from different locations in a text. He must synthesize ideas. Inferences directly implied within a text are included in this category. Also included are questions where the information is in two contiguous sentences and the reader must understand the role of sentence order in implying connections (see example 1 below). In other questions pronouns or general words refer to elements elsewhere in the text and the reader must make the appropriate referential connections in order to integrate the material (see example 2 below).

Samples:

> Why did Peter hide behind the stairs?
>
> (Answer: so he could surprise Willie; type - relationship
>
> What trick did Peter want to do with his dog?
>
> (Answer: whistle to call Willie; type - event detail)

Pilot Data:

Number of Items	11
Number of Students	73
Mean Total Score	6.11
Standard Deviation	2.22
Mean Proportion Correct	.56
Internal Consistency	.55

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	2	3	4	ROW TOTAL
1 LOW STRATA	4.545 11	0.0 0	6.000 4	5.667 6	5.143 21
2 MIDDLE STRATA	5.286 7	5.000 2	4.875 8	6.571 7	5.500 24
3 HIGH STRATA	6.600 5	6.500 4	8.500 8	7.000 3	7.400 20
COLUMN TOTAL	5.217 23	6.000 6	6.550 20	6.313 16	5.969 65

Revisions: None

T4. Comprehension of Text - Inferential, Extrapolation

These questions require the reader to relate the text to his own experience and knowledge. He must supply from experience something not present in the text. The most frequent kind of question is one which requires the reader to predict what might happen next, to go beyond the information given. (Questions of this type are the most open to alternative interpretations. Care was taken to try to write questions which would relate to common experiences.)

Samples:

> How do you think Peter felt at the end of the story?
>
> (Answer: happy; type - description (character))
>
> What lesson can we learn from this story?
>
> (Answer: keep trying and you will learn to do something (If at first you don't succeed.....); type - main idea)

Pilot Data:

Number of Items	9
Number of Students	73
Mean Total Score	3.77
Standard Deviation	2.14
Mean Proportion Correct	.42
Internal Consistency	.63

Ability-OTL Matrix:

MEAN / COUNT	0	2	3	4	ROW TOTAL
VOCSTRAT					
1 LOW STRATA	1.727 11	0.0 0	2.500 4	3.167 6	2.286 21
2 MIDDLE STRATA	3.286 7	3.714 7	1.333 3	4.286 7	3.458 24
3 HIGH STRATA	4.600 5	5.400 10	4.000 2	4.333 3	5.200 20
COLUMN TOTAL	2.826 23	4.706 17	2.444 9	4.250 16	3.615 65

Revisions: None

T5. Comprehension of Text - Evaluation

An additional dimension of comprehension comes into play at grade 5 - evaluation of information. Evaluation can be regarded as an additional category in the processing dimension. It is not crossed with the type of information dimension because it is difficult to find questions of each type which are not open to interpretation. The aspects of evaluation tested here are the ability to differentiate fact from opinion and the ability to judge the qualifications of a speaker.

Sample:

> Which of the following is a fact rather than an opinion?
>
> A. The Etruscans built cities to honor the dead.
>
> B. Historians do not know as much as archaeologists do.
>
> C. The Etruscans had a very clever way of hunting for food.
>
> D.
>
> E.

Pilot Data:

Number of Items	6
Number of Students	73
Mean Total Score	2.32
Standard Deviation	1.40
Mean Proportion Correct	.39
Internal Consistency	.36

Ability - OTL Matrix:

MEAN COUNT / VOCSTRAT	0	1	2	3	4	ROW TOTAL
1 LOW STRATA	1.364 11	0.0 0	1.444 9	0.0 0	2.000 1	1.429 21
2 MIDDLE STRATA	1.714 7	3.000 5	2.600 5	2.000 2	2.800 5	2.417 24
3 HIGH STRATA	2.000 5	2.714 7	2.500 2	4.333 3	5.333 3	3.150 20
COLUMN TOTAL	1.609 23	2.833 12	1.938 16	3.400 5	3.556 9	2.323 65

Revisions: None

III A. Comprehension - Applied

This scale tests the reader's ability to deal with material other than straight prose. Many of the same cognitive skills of comprehension are probably involved, but the reader must apply his skills to a specific task or format. Many of the items require the reader to search through or scan the page to locate information. Other items test the reader's orientation to the material, his understanding of what kind of information is being presented. The following components are included in the scale (with a separate OTL matrix for each):

	Group A	Group B
1. Letters		1 stem - 2 ?'s
2. Signs	2 items	2 items
3. Directions	1 stem - 2 ?'s	1 stem - 3 ?'s
4. Rules		2 stems - 6 ?'s
5. Maps	1 stem - 4 ?'s	
6. Graphs	3 stems - 8 ?'s	

Pilot Data:

	Group A	Group B
Number of Items	16	13
Number of Students	80	73
Mean Total Score	11.31	8.75
Standard Deviation	3.45	2.52
Mean Proportion Correct	.71	.67
Internal Consistency	.79	.66

1. Letters

Ability-OTL Matrix:

MEAN / COUNT	0	1	2	ROW TOTAL
VOCSTRAT				
1 LOW STRATA	1.091 / 11	1.300 / 10	0.0 / 0	1.100 / 21
2 MIDDLE STRATA	1.714 / 7	1.500 / 12	1.200 / 5	1.500 / 24
3 HIGH STRATA	1.200 / 5	1.375 / 8	1.857 / 7	1.700 / 20
COLUMN TOTAL	1.304 / 23	1.533 / 30	1.583 / 12	1.462 / 65

Revisions: None

2. Signs

Ability-OTL Matrix (Group A):

MEAN COUNT / VOCSTRAT	0	1	2	ROW TOTAL
1 LOW STRATA	0.588 17	0.000 5	2.000 1	0.652 23
2 MIDDLE STRATA	1.143 7	1.500 12	2.000 7	1.538 26
3 HIGH STRATA	1.667 3	1.000 16	1.400 5	1.625 24
COLUMN TOTAL	0.852 27	1.455 33	1.769 13	1.288 73

Ability-OTL Matrix (Group B):

MEAN COUNT / VOCSTRAT	0	1	2	ROW TOTAL
1 LOW STRATA	0.909 11	1.000 5	1.000 5	0.952 21
2 MIDDLE STRATA	1.857 7	1.800 10	1.429 7	1.708 24
3 HIGH STRATA	1.000 5	1.875 8	1.429 7	1.650 20
COLUMN TOTAL	1.348 23	1.652 23	1.316 19	1.446 65

Revisions: The two sign items in group B were eliminated (one was a traffic sign, the other a coupon). They appeared to be easy items, even without instruction. The group A items, both based on a set of bulletin board notices, were retained.

3. Directions

Ability-OTL Matrix (Group A):

MEAN COUNT / VOCSTRAT	0	1	2	3	ROW TOTAL
1 LOW STRATA	1.176 17	1.000 1	0.0 0	0.800 5	1.087 23
2 MIDDLE STRATA	1.143 7	1.500 2	2.000 3	1.571 14	1.500 26
3 HIGH STRATA	2.000 3	0.0 0	1.667 9	1.917 12	1.833 24
COLUMN TOTAL	1.259 27	1.333 3	1.750 12	1.581 31	1.479 73

Ability-OTL Matrix (Group B):

MEAN COUNT / VOCSTRAT	0	1	2	3	ROW TOTAL
1 LOW STRATA	2.000 11	3.000 5	0.0 0	2.400 5	2.333 21
2 MIDDLE STRATA	2.857 7	2.500 2	2.667 3	2.500 12	2.625 24
3 HIGH STRATA	2.000 5	0.0 0	2.857 7	3.000 8	2.700 20
COLUMN TOTAL	2.261 23	2.857 7	2.800 10	2.640 25	2.554 65

Revisions: None

4. Rules

Ability-OTL Matrix:

MEAN COUNT / VOCSTRAT	0	1	2	ROW TOTAL
1 LOW STRATA	2.182 11	2.400 5	2.800 5	2.381 21
2 MIDDLE STRATA	3.429 7	3.000 6	3.364 11	3.292 24
3 HIGH STRATA	3.200 5	4.000 3	4.417 12	4.050 20
COLUMN TOTAL	2.783 23	3.000 14	3.714 28	3.231 65

Revisions: None

6. Graphs

Ability-OTL Matrix:

MEAN CCUNT / VCCSTRAT	0	1	2	ROW TOTAL
1 LOW STRATA	4.600 5	4.900 10	4.500 8	4.696 23
2 MIDDLE STRATA	6.091 11	6.000 4	6.545 11	6.269 26
3 HIGH STRATA	6.800 10	5.333 6	7.000 8	6.500 24
CCLUMN TCTAL	6.077 26	5.250 20	6.074 27	5.849 73

Revisions: These items were dropped. Graphing is included in the math test, and was expanded in that test. The duplication of effort seemed unnecessary.

5. Maps

Ability-OTL Matrix (Reading):

MEAN COUNT / VOCSTRAT	0	1	2	3	ROW TOTAL
1 LOW STRATA	2.118 17	0.0 0	0.0 0	1.167 6	1.870 23
2 MIDDLE STRATA	2.571 7	3.000 2	4.000 1	2.563 16	2.654 26
3 HIGH STRATA	2.000 3	3.000 3	2.833 6	3.167 12	2.917 24
COLUMN TOTAL	2.222 27	3.000 5	3.000 7	2.529 34	2.493 73

Ability-OTL Matrix (Other Instruction):

MEAN COUNT / VOCSTRAT	0	1	4	ROW TOTAL
1 LOW STRATA	1.200 5	1.800 10	2.375 8	1.870 23
2 MIDDLE STRATA	2.636 11	2.750 4	2.636 11	2.654 26
3 HIGH STRATA	3.200 10	2.667 6	2.750 8	2.917 24
COLUMN TOTAL	2.577 26	2.250 20	2.593 27	2.493 73

Revisions: None

R. Reference Sources

1. Reference book
2. Reference source within a book

By grade 5, reading is an important means of gaining information. Students are expected to use their reading skill to study and learn new information. A study skill related to this use of reading is the ability to choose an appropriate source of information. A student with a particular problem or question needs to know what reference book will supply the answer. His experience with reading different reference sources should build a knowledge of the characteristics of different sources. Subtest R1 tests the ability to select the appropriate reference source for a particular purpose.

Sample (R1):

You want to know how to pronounce the word gravity.

A) card catalog C) dictionary

B) encyclopedia D) map

Particular sections within books are also reference sources. The Table of Contents, Glossary, and Index can be used by a reader to guide and enhance his reading. Scale R2 tests the ability to choose the appropriate reference section within a book. Sample pages of each type (from a book on The History of Dance) are provided as part of the directions.

Sample (R2):

Which part of the book tells if this book has a chapter about folk dance?

A. the Table of Contents

B. the Glossary

C. the Index

R. Reference Sources Cont'd

Pilot Data (R1 + R2):

Number of Items	12
Number of Students	80
Mean Total Score	8.86
Standard Deviation	2.40
Mean Proportion Correct	.74
Internal Consistency	.71

Ability-OTL Matrix (R1):

MEAN COUNT / VOCSTRAT	0	1	2	3	4	ROW TOTAL
1 LOW STRATA	3.235 17	4.000 1	0.0 0	4.000 4	6.000 1	3.522 23
2 MIDDLE STRATA	4.429 7	6.000 2	5.667 3	5.364 11	6.000 3	5.269 26
3 HIGH STRATA	5.000 3	0.0 0	5.667 9	5.200 5	5.857 7	5.542 24
COLUMN TOTAL	3.741 27	5.333 3	5.667 12	5.050 20	5.909 11	4.808 73

Ability-OTL Matrix (R1 - Language):

MEAN COUNT / VOCSTRAT	0	1	4	ROW TOTAL
1 LOW STRATA	4.400 5	2.800 10	3.875 8	3.522 23
2 MIDDLE STRATA	5.636 11	4.750 4	5.091 11	5.269 26
3 HIGH STRATA	5.600 10	5.667 6	5.375 8	5.542 24
COLUMN TOTAL	5.385 26	4.050 20	4.815 27	4.808 73

R. Reference Sources Cont'd

Ability-OTL Matrix (R2 - Language):

VOCSTRAT	MEAN / COUNT	0	1	5	ROW TOTAL
1		2.600	2.700	2.750	2.696
LOW STRATA		5	10	8	23
2		4.364	3.500	4.364	4.231
MIDDLE STRATA		11	4	11	26
3		4.800	4.500	4.625	4.667
HIGH STRATA		10	6	8	24
COLUMN TOTAL		4.192	3.400	3.963	3.890
		26	20	27	73

Revisions: None

0. Organization

1. Topic Sentence
2. Relevant Detail

One of the most important skills in reading is the ability to identify the topic, theme, or main idea of a selection, to know what it is all about. In some types of prose paragraphs, a topic sentence states the main idea clearly, and the remaining sentences elaborate on the topic. A skilled reader recognizes this underlying structure. He understands the superordinate-subordinate relationship between the topic and the supporting details. Subtest 01 tests the readers' ability to select the topic sentence from a set of sentences about the same subject. Subtest 02 tests the understanding of relevant detail by asking the reader to identify the statement which does not belong with the topic sentence.

Sample (01):

Which sentence states the topic?

A. The grass has morning dew on it.

B. Sunrise is the best time of day.

C. The sky has a gorgeous color.

D. Birds begin chirping.

Sample (02): Which one does not belong?

Follow the rules for boating safety.

A. Wear a life preserver.

B. Lake Erie has many small boats.

C. Don't lean over the side of the boat

D. Too many people should not squeeze into the boat.

O. Organization Cont'd

		01.	02.
Pilot Data:	Number of Items	5	5
	Number of Students	73	73
	Mean Total Score	3.04	3.62
	Standard Deviation	1.70	1.69
	Mean Proportion Correct	.61	.72
	Internal Consistency	.75	.82

Ability-OTL Matrix (01 + 02):

MEAN COUNT / VOCSTRAT	0	1	2	3	ROW TOTAL
1 LOW STRATA	4.182 11	6.200 5	4.400 5	0.0 0	4.714 21
2 MIDDLE STRATA	8.000 7	5.500 2	7.222 9	5.333 6	6.833 24
3 HIGH STRATA	8.200 5	0.0 0	8.556 9	8.667 6	8.500 20
COLUMN TOTAL	6.217 23	6.000 7	7.130 23	7.000 12	6.662 65

Revisions: None

03. Organization - Outlining

The logical extension of subtests 01 and 02 is to organize the information in a passage according to outline format. In an outline the topic sentence or phrase is placed in a superordinate position and relevant details are subordinate to it. Subtest 03 tests the ability to organize information in an outline. The student must select the appropriate element to fit in a particular position in an outline.

Sample:

II. _____99_____

A. Used nets and traps

B. Used music
   1. Set out nets and traps
   2. Play melody on flute
   3. Animals go to music and traps

Which heading goes in blank 99?

a. How Etruscans play flutes
b. How Etruscan animals liked music
c. How Etruscans hunted for food
d. Etruscans used nets for food
e. Etruscans fished and hunted

Pilot Data:

Number of Items	6
Number of Students	73
Mean Total Score	2.29
Standard Deviation	1.59
Mean Proportion Correct	.38
Internal Consistency	.56

03. Organization - Outlining Cont'd

Ability-OTL Matrix:

MEAN COUNT	0	1	ROW TOTAL
VCCSTRAT			
1 LOW STRATA	1.545 11	1.700 10	1.619 21
2 MIDDLE STRATA	2.429 7	1.941 17	2.083 24
3 HIGH STRATA	1.800 5	3.400 15	3.000 20
COLUMN TOTAL	1.870 23	2.405 42	2.215 65

Revisions: This scale was dropped due to lack of coverage.

Grade 2 Mathematics

C1. Computation - Addition without renaming

This scale is designed to assess the pupil's knowledge of addition facts and the addition algorithm. The items range from basic 1-digit combinations to the addition of two 3-digit numbers. Items are presented in both horizontal and vertical form.

Sample:

11 + 5 = ☐	A. 16	C. 13
	B. 17	D. 14

Pilot Data:

Number of Items	20
Number of Students	63
Mean Total Score	17.10
Standard Deviation	2.87
Mean Proportion Correct	.85
Internal Consistency	.80

Ability-OTL Matrix:

MEAN / COUNT	1	3	4	5	ROW TOTAL
RAVSTRAT					
1 LOW STRATA	16.0 / 1	15.6 / 11	20.0 / 1	15.0 / 4	15.8 / 17
2 MIDDLE STRATA	16.0 / 1	16.5 / 11	18.4 / 5	19.0 / 2	17.2 / 19
3 HIGH STRATA	18.0 / 5	19.2 / 6	18.6 / 7	18.0 / 1	18.6 / 19
COLUMN TOTAL	17.4 / 7	16.7 / 28	18.6 / 13	16.6 / 7	17.2 / 55

Revisions: Greater use of alternate format.
e.g. ☐ - 5 = 11

C2. Computation - Addition with renaming

This scale is designed to assess the pupil's knowledge of addition facts and addition algorithm. The items range from the addition of a 1-digit number with a 2-digit number to the addition of two 2-digit numbers.

Two items are in the horizontal format and the remainder in the vertical format. Answer choices include distractors which reflect systematic errors. These are as follows:

1. Failure to rename ten's column. One's column renamed correctly
2. Inverse operation in either or both columns.

Sample:

57 + 13	A. 64 C. 70 B. 60 D. 44

Pilot Data:

Number of Items	8
Number of Students	63
Mean Total Score	5.19
Standard Deviation	2.65
Mean Porportion Correct	.65
Internal Consistency	.87

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	3	4	5	ROW TOTAL
1 LOW STRATA	3.0 4	4.0 3	4.0 10	3.8 17
2 MIDDLE STRATA	3.3 6	6.3 6	6.4 7	5.4 19
3 HIGH STRATA	0.0 0	6.8 13	6.2 6	6.6 19
COLUMN TOTAL	3.2 10	6.3 22	5.3 23	5.3 55

Revisions: Expanded to 10 items.

C3. Computation - Subtraction without renaming

This scale is designed to assess pupil's knowledge of basic facts and the subtraction algorithm. Items range from subtraction of two 1-digit numbers to the subtraction of two 3-digit numbers. Items are presented in both a horizontal and vertical format.

Sample:

39 - 5 = ☐	A. 23	C. 33
	B. 24	D. 34

Pilot Data:

Number of Items	20
Number of Students	63
Mean Total Score	13.51
Standard Deviation	5.25
Mean Proportion Correct	.68
Internal Consistency	.92

Ability-OTL Matrix:

MEAN / COUNT	2	3	4	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	14.0 / 5	16.2 / 5	8.7 / 7	12.5 / 17
2 MIDDLE STRATA	16.3 / 3	19.4 / 5	9.4 / 11	13.1 / 19
3 HIGH STRATA	13.8 / 6	18.6 / 5	15.1 / 8	15.6 / 19
COLUMN TOTAL	14.4 / 14	18.1 / 15	11.0 / 26	13.8 / 55

Revisions: Greater use of alternate formats.
e.g. ☐ + 5 = 39

C4. Computation - Subtraction with renaming

This scale is designed to assess the pupil's knowledge of basic facts and the subtraction algorithm. Items range from the subtraction of a 1-digit from a 2-digit number to the subtraction of two 2-digit numbers. Items are presented in both a vertical and horizontal format. Four answer choices are provided including distractors representing systematic errors. These are as follows:

1. Failure to rename minuend. Subtracted smaller digit in minuend from subtrahend.
2. Subtracted minuend from subtrahend in one's column and "borrowed" from ten's column.
3. Failure to rename 10's column. One's column renamed correctly.

Sample:

94 - 75	A. 21 C. 29 B. 19 D. 11

Pilot Data:

Number of Items	8
Number of Students	63
Mean Total Score	2.08
Standard Deviation	2.63
Mean Proportion Correct	.33
Internal Consistency	.90

Ability-OTL Matrix:

MEAN / COUNT	3	4	5	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	0.0 4	0.8 4	2.1 9	1.3 17
2 MIDDLE STRATA	0.7 6	1.0 6	4.3 7	2.1 19
3 HIGH STRATA	0.0 0	1.6 13	5.5 6	2.8 19
COLUMN TOTAL	0.4 10	1.3 23	3.7 22	2.1 55

Revisions: Expanded to 10 items.

C5. Computation - Multiplication

This scale is designed to assess the pupil's knowledge of multiplication facts. Eight of the items involve basic 1-digit combinations with neither factor being greater than 5. Six items involve 1-digit combinations with one factor being $>5<9$ and the other factor being ≤ 5. Items are presented in both a vertical and horizontal format and include a distractor reflecting operational error of adding instead of multiplying.

Sample:

4 x 3 = ☐	A. 7 C. 11 B. 12 D. 15

Pilot Data:

Number of Items	14
Number of Students	63
Mean Total Score	7.52
Standard Deviation	4.98
Mean Proportion Correct	.54
Internal Consistency	.93

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	2.9 13	0.0 0	9.0 4	4.4 17
2 MIDDLE STRATA	5.9 12	14.0 1	12.3 6	8.4 19
3 HIGH STRATA	5.0 4	14.0 2	9.2 13	8.8 19
COLUMN TOTAL	4.4 29	14.0 3	10.0 23	7.3 55

Revisions: Reduced number of items to 10.

C6. Computation - Automaticity - Addition (1-1/2 minutes)

This scale is designed to assess the pupil's ability to recall basic addition facts rapidly. The first twelve items range from the addition of two 1-digit numbers to the addition of two 2-digit numbers with no renaming. The last two items require renaming. All items are presented in the vertical format. OTL matrix combines C6 and C7.

Sample:

6 + 5	A. 12 C. 11 B. 1 D. 9

Pilot Data:

Number of Items	14
Number of Students	63
Mean Total Score	11.03
Standard Deviation	3.38
Mean Proportion Correct	.79
Internal Consistency	.87

Ability-OTL Matrix (C6 & C7):

MEAN / COUNT		1	5	ROW TOTAL
RAVSTRAT				
LOW STRATA	1	13.8 5	22.1 12	19.6 17
MIDDLE STRATA	2	17.2 11	25.6 8	20.7 19
HIGH STRATA	3	20.4 7	22.9 12	22.0 19
COLUMN TOTAL		17.4 23	23.3 32	20.8 55

Revisions: Reduce time limit to one minute.

C7. Computation - Automaticity - Subtraction (2 minutes)

This scale is designed to assess the pupil's ability to recall basic subtraction facts rapidly. Items range from the subtraction of two 1-digit numbers to the subtraction of two 2-digit numbers. Only the last item requires renaming. All items are presented in the vertical format.

Sample:

18 - 9	A. 8	C. 11
	B. 9	D. 7

Pilot Data:

Number of Items	14
Number of Students	63
Mean Total Score	9.21
Standard Deviation	3.56
Mean Proportion Correct	.66
Internal Consistency	.86

Ability-OTL Matrix:

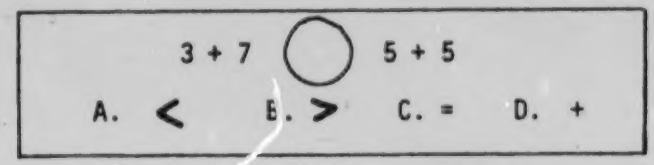
(See C6)

Revisions: None.

C8. Number Sentences - Equalities and Inequalities

This scale is designed to assess the pupil's ability to complete number sentences by selecting the appropriate equality, inequality or operational symbol (<, >, =, +, -).

Sample:



Pilot Data:

Number of Items	10
Number of Students	63
Mean Total Score	4.52
Standard Deviation	2.71
Mean Proportion Correct	.45
Internal Consistency	.76

Ability-OTL Matrix:

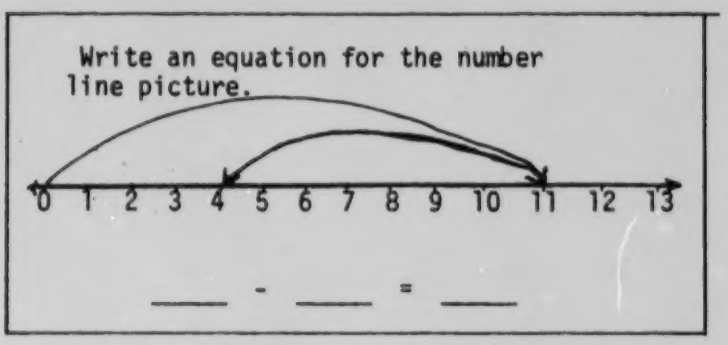
RAVSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	2.8 4	4.2 5	2.3 3	3.0 5	3.2 17
2 MIDDLE STRATA	6.0 2	7.0 5	6.0 1	4.3 11	5.3 19
3 HIGH STRATA	5.0 1	7.4 5	4.0 6	4.0 7	4.9 19
COLUMN TOTAL	4.0 7	6.2 15	3.7 10	3.9 23	4.5 55

Revisions: None

C9. Number line

This scale is designed to test the pupil's ability to interpret a numberical system as it is geometrically represented by the number line. Two items require the completion of an equation to match the given # line picture. Two items require the writing of an equation to match the picture. These four items are free response items. The remaining items are multiple choice and involve the recognition and selection of the equation matching the given # line picture.

Sample:



Pilot Data:

Number of Items	6
Number of Students	63
Mean Total Score	1.67
Mean Proportion Correct	.28

Due to the format this data had to be hand scored and was analyzed differently than other scale data. Therefore, data on internal consistency and standard deviation is not available.

Ability-OTL Matrix:

MEAN / COUNT	RAVSTRAT	1	2	ROW TOTAL
1	LOW STRATA	1.7 / 9	1.1 / 8	1.4 / 17
2	MIDDLE STRATA	3.0 / 7	2.5 / 12	2.7 / 19
3	HIGH STRATA	3.3 / 9	2.5 / 10	2.9 / 19
COLUMN TOTAL		2.6 / 25	2.1 / 30	2.4 / 55

60. (continued)

Revisions: This scale has been deleted from the revised test battery. Both teacher and consultant feedback, as well as data results, indicate that this scale is testing the pupil's familiarity with a specific instructional approach. Familiarity with different methods of pictorial representation does not seem to generalize across methods.

C10. Renaming Numerals (equation form). Correct Recognition.

This scale includes only those items (equations) which correctly rename a given numeral. There are four given numerals and six equations per numberal. This scale is intended to measure the pupil's knowledge of basic facts as well as the concept that each numberal has many names.

Sample:

> Circle all the equations that rename the numeral 7.
>
> A. 2 + 5 = ☐ D. 9 - 2 = ☐
>
> B. 6 - 1 = ☐ E. 10 - 4 = ☐
>
> C. 3 + 4 = ☐ F. 3 + 6 = ☐

Pilot Data:

Number of Items	14
Number of Cases	63
Mean Total Score	9.73
Standard Deviation	3.02
Mean Proportion Correct	.70
Internal Consistency	.79

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	5	ROW TOTAL
RAVSTRAT					
1 LOW STRATA	16.3 3	17.0 4	21.4 5	15.0 5	17.6 17
2 MIDDLE STRATA	16.0 1	17.5 2	22.2 5	15.5 11	17.5 19
3 HIGH STRATA	17.7 6	21.0 1	21.6 5	19.0 7	19.4 19
COLUMN TOTAL	17.1 10	17.7 7	21.7 15	16.4 23	18.2 55

Revisions: None.

C10(a). Renaming Numeral (equation form) - Incorrect recognition

This scale is extracted from the same scale of 24 equations used for correct recognition. It consists of 10 items (equations) which do not rename the given numeral. Scores are determined by the pupil's not identifying one of these equations as correct. The Ability-OTL Matrix combines C10 and C10a and is reported under C10.

Sample: See C10

Pilot Data:

Number of Items	10
Number of Students	63
Mean Total Score	8.27
Standard Deviation	2.28
Mean Proportion Correct	.83
Internal Consistency	.81

Ability-OTL Matrix:

(see C10)

Revisions: None

N1,2,3. Numerals/Ordinals/Place Value

This scale tests the pupil's knowledge of three components of the number system. It includes four items on numerals, four on ordinals, and nine on place value using both compact and expanded notation. These subtests were treated as one scale for the item analysis but were considered separately for the Ability-OTL Matrix data.
Pilot data is reported for the scale as a whole and samples, OTL Matrix, etc., are given for each subtest.

Pilot Data:

Number of Items	17
Number of Students	63
Mean Total Score	11.49
Standard Deviation	3.86
Mean Proportion Correct	.68
Internal Consistency	.82

Sample N1:

> Which of these numerals names the smallest number?
>
> A. 390 B. 379 C. 399 D. 400

Number of items	4
Mean Proportion Correct.	.72

Ability-OTL Matrix:

RAWSTRAT	MEAN / COUNT	5	ROW TOTAL
1 LOW STRATA		2.3 17	2.3 17
2 MIDDLE STRATA		2.8 19	2.8 19
3 HIGH STRATA		3.5 19	3.5 19
COLUMN TOTAL		2.9 55	2.9 55

N1,2,3. (page 2)

Sample N2. Ordinals:

> Which letter is third letter in the word, FROG?
>
> A) O B) R C) G D) F

Number of Items 4
Mean Proportion Correct. .79

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	4	ROW TOTAL
1 LOW STRATA	2.8 12	0.0 0	2.8 5	2.8 17
2 MIDDLE STRATA	3.6 7	4.0 1	2.8 11	3.2 19
3 HIGH STRATA	3.7 10	4.0 2	3.6 7	3.7 19
COLUMN TOTAL	3.3 29	4.0 3	3.0 23	3.2 55

N1,2,3. (page 3)

Sample N3. Place Value:

What is another name for 59?	
A) 5 + 9	C) 9 + 5
B) 90 + 5	D) 50 + 9

Number of items 9
Mean proportion correct .61

Ability-OTL Matrix:

MEAN / COUNT	3	4	5	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	2.5 / 2	5.3 / 6	5.0 / 9	4.8 / 17
2 MIDDLE STRATA	8.0 / 1	6.2 / 5	4.6 / 13	5.2 / 19
3 HIGH STRATA	7.0 / 6	6.0 / 5	6.9 / 8	6.7 / 19
COLUMN TOTAL	6.1 / 9	5.8 / 16	5.3 / 30	5.6 / 55

Revisions: The subtests on numerals and ordinals have been combined and expanded by 2 items to form a 10 item scale. Place value has been expanded to form a 15 item scale. New items include pictorial representations and 5 open-ended items.

N4(a). Fractions Involving Sets

This scale is an extension of C7 and involves the identification of a fractional part of a set.

Sample:

> The set ✱✱ is $\frac{2}{3}$ of which set below?
>
> A. ✱✱✱✱ C. ✱✱✱✱✱
>
> B. ✱✱ D. ✱ ✱✱

Pilot Data:

Number of Items	5
Number of Students	78
Mean Total Score	1.55
Standard Deviation	1.01
Mean Proportion Correct	.31
Internal Consistency	.30

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	0.5 2	1.3 8	1.8 9	1.5 6	1.4 25
2 MIDDLE STRATA	1.0 1	1.5 7	1.2 10	1.3 6	1.3 24
3 HIGH STRATA	0.5 2	1.7 10	1.9 9	2.3 4	1.8 25
COLUMN TOTAL	0.6 5	1.5 25	1.6 28	1.6 16	1.5 74

Revisions: Two items were dropped and replaced by open ended items. The total scale (C7 and C8) was expanded to 15 items to include lines and fractional relationships.

N4(b). Fractions Involving Regions

This scale tests the pupil's ability to identify the fractional part of a region.

Sample:

What part of this region has been shaded?

A. $\frac{1}{2}$ B. $\frac{1}{4}$ C. $\frac{1}{3}$ D. $\frac{2}{3}$

Pilot Data:

Number of Items	5
Number of Students	78
Mean Total Score	2.59
Standard Deviation	1.53
Mean Proportion Correct	.51
Internal Consistency	.70

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	1.0	2.0	2.9	2.0	2.2
	2	8	9	6	25
2 MIDDLE STRATA	1.0	2.2	2.8	2.3	2.5
	1	5	12	6	24
3 HIGH STRATA	1.5	1.7	3.5	3.0	2.8
	2	6	13	4	25
COLUMN TOTAL	1.2	1.9	3.1	2.4	2.5
	5	19	34	16	74

Revisions: None

N5. Odd and Even Numbers

This scale tests the pupils knowledge of odd and even numbers.

Sample:

> Which number is an even number?
>
> A) 7 B) 9 C) 11 D) 12

Pilot Data:

Number of Items	4
Number of Students	78
Mean Total Score	1.69
Standard Deviation	1.42
Mean Proportion Correct	.42
Internal Consistency	.70

Ability-OTL Matrix:

MEAN / COUNT	1	2	ROW TOTAL
RAVSTRAT			
1 LOW STRATA	1.1 / 13	1.1 / 12	1.1 / 25
2 MIDDLE STRATA	1.3 / 12	1.6 / 12	1.5 / 24
3 HIGH STRATA	2.7 / 16	1.9 / 9	2.4 / 25
COLUMN TOTAL	1.8 / 41	1.5 / 33	1.6 / 74

Revisions: This scale has been dropped since classroom instruction appears to be minimal for this objective.

P1. Properties

This scale is designed to measure a pupil's knowledge of the associative and commutative properties in respect to the addition and multiplication of whole numbers. The first 6 items conform to the standard format for testing properties (sample Item A, B and C). The last 4 items combine the associative property with expanded notation (sample Item D). The latter is an attempt to tap the pupil's understanding of this combination which is used in the teaching of addition without regrouping.

The item analysis data combined these ten items into one scale and the Ability-OTL Matrix was done separately for each property. Therefore, pilot data is reported for the whole scale and the sample items, OTL-Matrix, number of items and mean proportion correct are given for each subscale.

Pilot Data:

Number of Items	10
Number of Students	63
Mean Total Score	3.90
Standard Deviation	2.67
Mean Proportion Correct	.39
Internal Consistency	.75

Sample A: Associative

Which numeral goes in the box?

(20 + 3) + 7 = ☐ + (3 + 7)

A. 10 B. 14 C. 23 D. 20

Number of Items	2
Mean Proportion Correct	.34

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	0.3 4	0.4 13	0.0 0	0.4 17
2 MIDDLE STRATA	0.0 2	0.7 16	2.0 1	0.7 19
3 HIGH STRATA	0.0 1	1.0 16	1.0 2	0.9 19
COLUMN TOTAL	0.1 7	0.7 45	1.3 3	0.7 55

P1. Page two.

Sample B: Commutative (+)

16 + 77 = ☐ + 16

A. 77 B. 32 C. 61 D. 16

Number of items 2
Mean Proportion Correct .40

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	2	3	ROW TOTAL
1 LOW STRATA	0.3 9	0.8 8	0.5 17
2 MIDDLE STRATA	1.4 7	0.8 12	1.1 19
3 HIGH STRATA	0.5 6	1.1 13	0.9 19
COLUMN TOTAL	0.7 22	0.9 33	0.8 55

Sample C: Commutative (x)

4 x 2 = 2 x ☐

A. 4 B. 2 C. 8 D. 3

Number of items 2
Mean Proportion Correct .33

P1. Page three.

Ability-OTL Matrix:

RAVSTRAT	MEAN COUNT	1	2	4	ROW TOTAL
1 LOW STRATA		0.2	0.5	1.0	0.4
		11	4	2	17
2 MIDDLE STRATA		1.0	0.9	0.7	0.8
		6	7	6	19
3 HIGH STRATA		0.2	1.0	1.3	0.9
		5	2	12	19
COLUMN TOTAL		0.4	0.8	1.0	0.7
		22	13	20	55

Sample D: Associative With Expanded Notation

> Another way to write 17 + 3 is
>
> A. 10 + 7 (7+3) B. 1 + (7x3) C. 10 + (70+3)

Number of items 4
Mean Proportion Correct .44

Ability-OTL Matrix (Using Associative Property OTL):

RAVSTRAT	MEAN COUNT	1	2	3	ROW TOTAL
1 LOW STRATA		1.3	1.9	0.0	1.8
		4	13	0	17
2 MIDDLE STRATA		1.5	1.6	4.0	1.7
		2	16	1	19
3 HIGH STRATA		1.0	1.9	4.0	2.1
			16	2	19
COLUMN TOTAL		1.3	1.8	4.0	1.9
		7	45	3	55

P1. Page four.

Ability-OTL Matrix (Using Place Value OTL):

MEAN COUNT / RAVSTRAT	3	4	5	ROW TOTAL
1 LOW STRATA	1.5 2	2.0 6	1.7 9	1.8 17
2 MIDDLE STRATA	4.0 1	1.8 5	1.5 13	1.7 19
3 HIGH STRATA	2.7 6	1.4 5	2.0 8	2.1 19
COLUMN TOTAL	2.6 9	1.8 16	1.7 30	1.9 55

Revisions: Expanded by two items to include the identity principle.

M1. Money.

This scale tests the pupil's knowledge of coins and his/her ability to identify thie sum of a set of coins.

Sample:

What is the value of these coins?

A) 9¢

B) 14¢

C) 45¢

D) 41¢

Pilot Data:

Number of Items	5
Number of Students	78
Mean Total Score	3.88
Standard Deviation	1.22
Mean Proportion Correct	.78
Internal Consistency	.63

Ability-OTL Matrix:

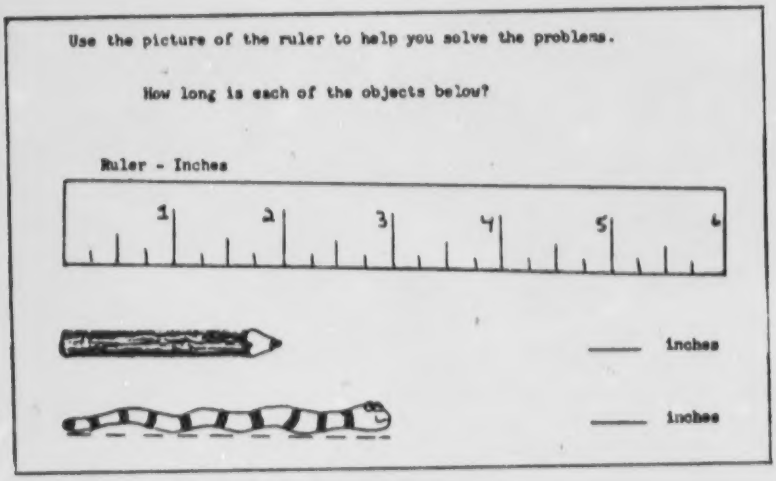
MEAN COUNT / RAVSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	1.5 2	4.0 1	3.6 16	3.3 6	3.4 25
2 MIDDLE STRATA	5.0 1	3.4 5	4.5 12	3.3 6	4.0 24
3 HIGH STRATA	4.5 2	3.9 10	4.6 9	4.5 4	4.3 25
COLUMN TOTAL	3.4 5	3.8 16	4.1 37	3.6 16	3.9 74

Revisions: One item dropped and four items added to form an 8 item scale. New items use coins in an equational format and are open ended.

M2. Linear Measurement

This scale tests the pupil's ability to estimate the length or height of an object when both the object and ruler are presented pictorially. All seven items are open ended.

Sample:



Pilot Data:

Number of Items	7
Number of Students	78
Mean Total Score	4.51
Standard Deviation	1.45
Mean Proportion Correct	.64
Internal Consistency	.66

Ability-OTL Matrix:

MEAN COUNT / PAVSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	3.5 2	3.8 18	5.2 5	4.0 25
2 MIDDLE STRATA	5.0 1	4.4 18	5.6 5	4.7 24
3 HIGH STRATA	5.5 2	4.9 15	4.4 8	4.8 25
COLUMN TOTAL	4.6 5	4.3 51	4.9 18	4.5 74

Revisions: None

M3. Measurement Concepts

This scale is designed to measure the pupil's ability to recognize the attributes of objects which can or cannot be measured. A scene/ picture is given along with six statements (Sample A) which can be circled if that attribute can be measured.

An additional 4 items (Sample B) test pre-measurement concepts of order and height differences. Pilot data is reported for each of these "subtests" and OTL data is given for the total scale.

Sample A: (Partial)

A. How tall Maria is.

B. How much water is in the bowl.

Pilot Data:

Number of Items	6
Number of Students	78
Mean Total Score	4.46
Standard Deviation	1.42
Mean Proportion Correct	.74
Internal Consistency	.54

Sample B:

Which set of objects goes from the largest to the smallest?

A.

C.

B.

D.

Pilot Data:

Number of Items	4
Number of Students	78
Mean Total Score	3.33
Standard Deviation	1.06
Mean Proportion Correct	.84
Internal Consistency	.72

M3. Page two.

Ability-OTL Matrix:

MEAN COUNT	3	ROW TOTAL
RAVSTRAT		
1 LOW STRATA	7.0 25	7.0 25
2 MIDDLE STRATA	7.9 24	7.9 24
3 HIGH STRATA	8.5 25	8.5 25
COLUMN TOTAL	7.8 74	7.8 74

Revisions: Pre-measurement items modified to increase difficulty.
Picture with statements dropped.

M4. Liquid Measurement

This scale tests the pupil's knowledge of the units of liquid measurement.

Sample:

How many pints are there in 1 quart?

A. 2 pints C. 4 pints

B. 3 pints D. 1 pint

Pilot Data:

Number of Items	5
Number of Students	78
Mean Total Score	1.92
Standard Deviation	1.25
Mean Proportion Correct	.38
Internal Consistency	.34

Ability-OTL Matrix:

RAVSTRAT	MEAN / COUNT	1	2	ROW TOTAL
1 LOW STRATA		1.5 / 13	1.6 / 12	1.5 / 25
2 MIDDLE STRATA		1.9 / 8	1.9 / 16	1.9 / 24
3 HIGH STRATA		2.0 / 3	2.4 / 22	2.3 / 25
COLUMN TOTAL		1.7 / 24	2.0 / 50	1.9 / 74

Revisions: Scale has been dropped since few teachers cover this objective.

M5. Time.

This scale tests the pupil's ability to recognize time to the hour, half-hour, and quarter hour.

Sample:

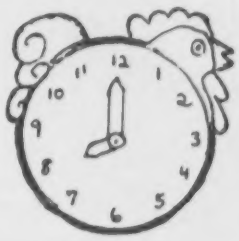
What time does the clock show?

A. 8 o'clock

B. Half-past 9

C. 12 o'clock

D. 10 o'clock



Pilot Data:

Number of Items	5
Number of Students	78
Mean Total Score	3.0
Standard Deviation	1.41
Mean Proportion Correct	.60
Internal Consistency	.62

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	1.0 / 2	3.5 / 6	2.1 / 17	2.3 / 25
2 MIDDLE STRATA	1.0 / 1	3.3 / 10	3.1 / 13	3.1 / 24
3 HIGH STRATA	3.0 / 2	3.4 / 18	3.6 / 5	3.4 / 25
COLUMN TOTAL	1.8 / 5	3.4 / 34	2.7 / 35	2.9 / 74

Revisions: This scale has been dropped as recognition of time is not really a component of mathematics. Two problems involving computation with time have been incorporated into the word problems scale.

G1,2. Geometry

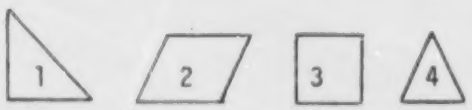
This scale tests the pupil's knowledge of geometric figures (5 items), open and closed curves (3 items), and points inside and outside closed curves (2 items). Pilot data is reported on this scale as a whole and OTL data is given separately for figures and curves/points.

Pilot Data:

Number of Items	10
Number of Students	78
Mean total score	7.62
Standard Deviation	2.01
Mean Proportion Correct	.76
Internal Consistency	.65

Sample A: Figures

Which of these geometric shapes are triangles?



A) 1 and 4 C) 1 and 2

B) 2 and 3 D) 3 and 4

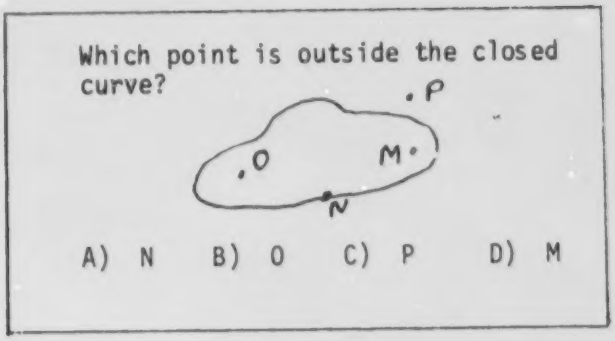
Number of Items	5
Mean Proportion Correct	.80

Ability- OTL Matrix:

MEAN COUNT / RAWSTRAT	2	3	4	ROW TOTAL
1 LOW STRATA	4.1 8	3.4 14	3.7 3	3.6 25
2 MIDDLE STRATA	4.1 7	3.8 11	4.3 6	4.0 24
3 HIGH STRATA	4.8 8	4.1 7	3.9 10	4.2 25
COLUMN TOTAL	4.3 23	3.7 32	4.0 19	4.0 74

G1,2. Page two.

Sample B: Points/Curves



Number of Items 5
Mean Proportion Correct .72

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	2.9 14	4.4 10	5.0 1	3.6 25
2 MIDDLE STRATA	3.3 11	4.0 10	3.7 3	3.6 24
3 HIGH STRATA	2.7 9	3.7 12	4.5 4	3.4 25
COLUMN TOTAL	3.0 34	4.0 32	4.3 8	3.6 74

Revisions: Total scale expanded to 14 items. Two new items on points, one on curves, and one on figures.

A1. Word Problems

This scale is designed to measure the pupil's ability to solve word problems. Eight of the items involve either addition or subtraction and the remaining two involve the joining of equivalent sets.

Sample:

> Mary baked 20 cookies. John and Lisa ate 12 of them. How many cookies are left?
>
> A. 32 B. 18 C. 8 D. 6

Pilot Data:

Number of Items	10
Number of Students	78
Mean Total Score	6.67
Standard Deviation	2.10
Mean Proportion Correct	.67
Internal Consistency	.61

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	4.0 3	5.8 18	7.8 4	0.0 0	5.9 25
2 MIDDLE STRATA	4.8 4	6.9 13	7.6 5	7.0 2	6.7 24
3 HIGH STRATA	5.4 8	8.0 5	7.8 8	8.5 4	7.2 25
COLUMN TOTAL	4.9 15	6.5 36	7.7 17	8.0 6	6.6 74

Revisions: Scale expanded to 15 items, 2 of which involve the partitioning of sets and 2 involve time.

01. Addition

This scale is designed to test the pupil's knowledge of basic facts and the addition algorithm. Imbedded within this scale are four items corresponding to the computation required by four of the word problems in scale A1. Six items involve regrouping and the remaining eight do not. Both the horizontal and vertical formats are used. For the OTL analysis this scale was separated into two subtests. One contains the 8 items that did not involve regrouping and the other has the 6 items with regrouping.

Sample:

16 + 15	A. 29 C. 31 B. 19 D. 21

Pilot Data:

Number of Items	14
Number of Students	78
Mean Total Score	10.68
Standard Deviation	3.27
Mean Proportion Correct	.76
Internal Consistency	.86

Ability-OTL Matrix (without regrouping):

MEAN COUNT / RAVSTRAT	2	3	4	5	ROW TOTAL
1 LOW STRATA	5.5 2	8.0 1	4.9 14	6.3 8	5.5 25
2 MIDDLE STRATA	4.0 1	7.7 3	7.1 13	7.0 7	7.0 24
3 HIGH STRATA	7.0 2	8.0 4	7.2 11	7.6 8	7.4 25
COLUMN TOTAL	5.8 5	7.9 8	6.3 38	7.0 23	6.6 74

01. Page two.

Ability-OTL Matrix (with regrouping):

MEAN COUNT / RAVSTRAT	2	3	4	5	ROW TOTAL
1 LOW STRATA	1.6	1.5	4.2	2.0	2.9
	7	2	12	4	25
2 MIDDLE STRATA	3.5	1.0	4.1	5.5	4.5
	2	1	10	11	24
3 HIGH STRATA	0.0	3.0	4.6	4.4	4.4
	0	2	5	18	25
COLUMN TOTAL	2.0	2.0	4.2	4.5	3.9
	9	5	27	33	74

Revisions: None

02. Subtraction

This scale is designed to assess the pupil's knowledge of basic facts and the subtraction algorithm. Imbedded within this scale are four items which correspond to four of the word problems in scale A1. Six items require regrouping, eight do not. Both the horizontal and vertical formats are used. The sample item below corresponds to the sample word problem given for scale A1. As in A2 the OTL analysis was done separately for the items with regrouping and those without.

Sample:

20 - 12	A. 18 C. 32 B. 8 D. 28

Pilot Data:

Number of Items	14
Number of Students	78
Mean Total Score	8.44
Standard Deviation	2.90
Mean Proportion Correct	.60
Internal Consistency	.78

Ability-OTL Matrix (without regrouping):

MEAN COUNT / RAVSTRAT	2	3	4	ROW TOTAL
1 LOW STRATA	8.0 1	5.7 11	5.0 13	5.4 25
2 MIDDLE STRATA	7.3 3	6.3 7	6.2 14	6.4 24
3 HIGH STRATA	6.3 4	7.0 1	7.0 20	6.9 25
COLUMN TOTAL	6.9 8	6.0 19	6.2 47	6.2 74

02. Page two.

Ability-OTL Matrix (with regrouping):

RAVSTRAT	MEAN COUNT	1	2	3	4	5	ROW TOTAL
1 LOW STRATA		1.4	1.5	1.3	2.4	3.0	1.9
		7	2	4	10	2	25
2 MIDDLE STRATA		2.0	3.0	3.0	0.7	3.4	2.0
		2	1	5	11	5	24
3 HIGH STRATA		0.0	2.5	1.0	2.5	2.0	2.3
		0	2	1	14	8	25
COLUMN TOTAL		1.6	2.2	2.1	1.9	2.6	2.1
		9	5	10	35	15	74

Revisions: None

Grade 5 Mathematics

C1. Addition and Subtraction of Whole Numbers

This scale is designed to test the pupil's ability to add and subtract whole numbers. The items involved 2 to 5 digit numerals, eight with regrouping and two without. Only nine items are being reported on here since one item went to print without a correct answer choice. Pilot data combines the addition and subtraction items into one scale. OTL data reports addition and subtraction as separate subtests. Distractors are included which reflect the following:

1. Addition - failed to "carry".
2. Subtraction - subtracted minuend from subtrahend, in one or more columns.
3. Subtracted minuend from subtrahend plus "borrowed" from 10's, 100's, etc., column.
4. Failed to rename 10's, 100's, etc. column.

Sample:

1934	A. 1079	C. 989
- 955	B. 1089	D. 979

Pilot Data:

Number of Items	9
Number of Students	79
Mean Total Score	7.63
Standard Deviation	1.81
Mean Proportion Correct	.85
Internal Consistency	.76

Ability-OTL Matrix (Addition):

MEAN / COUNT	2	3	5	ROW TOTAL
RAWSTRAT				
1 LOW STRATA	4.5 4	5.0 5	3.6 14	4.1 23
2 MIDDLE STRATA	5.0 2	4.7 9	4.4 13	4.5 21
3 HIGH STRATA	5.0 3	4.9 11	4.7 14	4.8 28
COLUMN TOTAL	4.8 9	4.9 22	4.2 41	4.5 72

C1. (continued)

Ability-OTL Matrix (Subtraction):

RAVSTRAT	MEAN COUNT	2	3	4	5	ROW TOTAL
1 LOW STRATA		3.5	3.0	1.5	2.6	2.6
		4	1	4	14	23
2 MIDDLE STRATA		4.0	0.0	3.0	2.7	2.9
		2	0	6	13	21
3 HIGH STRATA		4.0	3.8	3.2	3.4	3.5
		3	5	6	14	28
COLUMN TOTAL		3.8	3.7	2.7	2.9	3.0
		9	6	16	41	72

Revisions: Separated into two scales with 10 items each. Increased item difficulty. Refined distractors.

C2. Multiplication of Whole Numbers

This scale is designed to test the pupil's ability to multiply whole numbers.

Eight items involve "carrying" and two do not. Distractors reflecting the following errors are included.

1. Failure to carry - any or all columns
2. Error in position of partial products
3. Failure to carry in adding partial products

Sample:

193 x 79	A. 15,047 C. 15,027 B. 15,247 D. 3,088

Pilot Data:

Number of Items	10
Number of Student	79
Mean Total Score	7.59
Standard Deviation	2.55
Mean Proportion Correct	.76
Internal Consistency	.81

Ability-OTL Matrix:

MEAN / COUNT	2	4	5	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	7.0 3	8.3 6	5.8 14	6.6 23
2 MIDDLE STRATA	8.2 5	7.6 5	7.6 11	7.8 21
3 HIGH STRATA	7.8 4	8.7 15	8.3 9	8.4 28
COLUMN TOTAL	7.8 12	8.4 26	7.1 34	7.7 72

Revisions: Expanded scale to 15 items. Increased item difficulty. Refined distractors.

C3. Division of Whole Numbers

This scale is designed to test the pupil's ability to divide whole numbers. Items involve 1 and 2 digit divisors, 2 to 4 digit dividends, and quotients with and without remainders. Because of the large variety of possible errors in division, it is extremely difficult to systematically test for these errors using a multiple choice format. However, some types of distractors will be included as a possible sampling of errors. The number of items for which a particular error is appropriate is too small to make an error analysis profitable.

Sample:

336 ÷ 8	A. 23	C. 32
	B. 54	D. 42

Pilot Data:

Number of Items	10
Number of Students	79
Mean Total Score	4.16
Standard Deviation	2.45
Mean Proportion Correct	.41
Internal Consistency	.72

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	3	4	5	ROW TOTAL
1 LOW STRATA	1.0 3	4.0 1	4.2 11	2.9 8	3.3 23
2 MIDDLE STRATA	3.7 3	3.4 5	3.2 10	7.0 3	3.9 21
3 HIGH STRATA	1.0 1	5.0 5	5.5 15	4.6 7	5.0 28
COLUMN TOTAL	2.1 7	4.2 11	4.5 36	4.2 18	4.1 72

Revisions: Scale expanded to 15 items. Increased range of difficulty particularly at lower end. Dropped 4 items which were too difficult.

C4. Addition of Fractions

This scale is designed to test the pupil's ability to add two proper fractions with sums less than or equal to one. Three items involve like denominators and four have unlike denominators.

Sample:

$\frac{3}{7} + \frac{2}{7} = \square$	A. $\frac{1}{7}$ C. $\frac{7}{14}$ B. 1 D. $\frac{5}{7}$

Pilot Data:

Number of items	7
Number of Students	79
Mean Total Score	2.01
Standard Deviation	1.97
Mean Proportion Correct	.29
Internal Consistency	.81

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	1.3 17	1.0 1	3.0 5	1.7 23
2 MIDDLE STRATA	1.1 10	1.8 5	2.2 6	1.6 21
3 HIGH STRATA	0.8 10	1.9 10	5.3 8	2.5 28
COLUMN TOTAL	1.1 37	1.8 16	3.7 19	1.9 72

Revisions: Combined with C5 and C6 to form one 10 item scale. New scale includes 4 items from this test. Revision reflects decision to increase emphasis on whole numbers while reducing the emphasis on fractions to better reflect classroom instruction. Distractors are listed under C5.

C5. Addition of Fractions

This scale is an extension of C4 and tests the pupil's ability to add two proper fractions with like denominators whose sum is greater than one. All sums are in mixed numeral form.

Sample:

$\frac{5}{8} + \frac{6}{8} = \square$	A. $\frac{1}{8}$ C. $\frac{3}{8}$ B. $1\frac{1}{8}$ D. $1\frac{3}{8}$

Pilot Data:

Number of Items	7
Number of Students	79
Mean Total Score	1.84
Standard Deviation	2.00
Mean Proportion Correct	.26
Internal Consistency	.79

Ability-OTL Matrix:

MEAN COUNT		1	2	3	ROW TOTAL
RAVSTRAT					
LOW STRATA	1	1.0 17	1.0 1	2.2 5	1.3 23
MIDDLE STRATA	2	1.6 10	1.0 5	2.2 6	1.6 21
HIGH STRATA	3	0.7 10	1.0 10	4.9 8	2.0 28
COLUMN TOTAL		1.1 37	1.0 16	3.3 19	1.7 72

Revisions: See C4. 3 Items from this scale are included in new scale. Distractors include the following:

1. Added denominators and numerators
2. Denominators added to get common denominator
3. Multiplied numerators and added denominators.

C6. Addition of Fractions

This scale is an extension of C4 and C5 and involves the addition of a mixed number and a proper fraction or two mixed numbers. Two items have unlike denominators.

Sample:

$5\frac{3}{8}$ $+\ 6\frac{1}{4}$	A. $10\frac{5}{8}$ C. $11\frac{1}{2}$ B. $11\frac{5}{8}$ D. $11\frac{6}{8}$

Pilot Data:

Number of Items	7
Number of Students	79
Mean Total Score	2.35
Standard Deviation	2.10
Mean Proportion Correct	.34
Internal Consistency	.80

Ability-OTL Matrix:

MEAN / COUNT	1	2	4	ROW TOTAL
RAWSTRAT				
1 LOW STRATA	1.2 18	0.5 4	0.0 1	1.0 23
2 MIDDLE STRATA	2.0 13	3.0 6	0.0 0	2.6 21
3 HIGH STRATA	2.2 17	2.7 6	6.4 5	3.0 28
COLUMN TOTAL	1.8 48	2.6 16	5.3 6	2.3 72

Revisions: See C4 and C5. Three items were kept from this scale.

C7. Subtraction with Fractions

This scale tests the pupil's ability to subtract a proper fraction from a proper fraction or whole number. Three items involve fractions with unlike denominators.

Sample:

$\frac{3}{6}$	A. $\frac{1}{4}$	C. $\frac{5}{12}$
$-\frac{1}{12}$	B. $\frac{2}{12}$	D. $\frac{1}{2}$

Pilot Data:

Number of Items	7
Number of Students	79
Mean Total Score	1.77
Standard Deviation	1.61
Mean Proportion Correct	.25
Internal Consistency	.64

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	ROW TOTAL
1 LOW STRATA	0.9 23	0.0 0	0.9 23
2 MIDDLE STRATA	1.6 19	2.5 2	1.7 21
3 HIGH STRATA	2.4 25	2.3 3	2.4 28
COLUMN TOTAL	1.7 67	2.4 5	1.7 72

Revisions: Combined with C8 to form a new scale of 10 items, five of which are from this scale and 5 are from C8. Distractors include:

1. Subtracted numerators and added denominators
2. Subtracted numerators and multiplied denominators
3. Failed to rename whole number.

C8. Subtraction With Fractions

This is an extension of C7 and involves the subtraction of a proper fraction or mixed number from a mixed number. Four items have unlike denominators.

Sample:

$8\ \frac{7}{9} - 3\ \frac{2}{9} = \square$	A. $6\ \frac{5}{9}$ C. $6\ \frac{4}{9}$ B. $5\ \frac{1}{3}$ D. $5\ \frac{5}{9}$

Pilot Data:

Number of Items	7
Number of Students	79
Mean Total Score	2.08
Standard Deviation	2.14
Mean Proportion Correct	.30
Internal Consistency	.81

Ability-OTL Matrix:

MEAN / COUNT	1	2	ROW TOTAL
RAWSTRAT			
1 LOW STRATA	1.0 23	0.0 0	1.0 23
2 MIDDLE STRATA	1.9 19	1.5 2	1.9 21
3 HIGH STRATA	2.7 25	2.3 3	2.7 28
COLUMN TOTAL	1.9 67	2.0 5	1.9 72

Revisions: See C7.

C9. Multiplication of Fractions

This scale tests the pupil's ability to multiply a proper fraction by a whole number or another proper fraction. Three items involve fractions with unlike denominators.

Sample:

$\frac{1}{2} \times \frac{1}{3} = \square$	A. $\frac{1}{2}$ C. $\frac{1}{5}$ B. $\frac{1}{3}$ D. $\frac{1}{6}$

Pilot Data:

Number of Items	7
Number of Students	79
Mean Total Score	1.78
Standard Deviation	1.57
Mean Proportion Correct	.25
Internal Consistency	.60

Ability-OTL Matrix:

MEAN / COUNT	1	2	4	ROW TOTAL
RAWSTRAT				
1 LOW STRATA	1.6 21	4.0 1	2.0 1	1.7 23
2 MIDDLE STRATA	1.5 16	2.0 5	0.0 0	1.7 21
3 HIGH STRATA	1.6 13	1.5 10	3.0 5	1.8 28
COLUMN TOTAL	1.6 50	1.9 16	2.8 6	1.9 72

Revisions: This scale was dropped along with C10 and has been replaced by a 10 item scale on equivalent fractions. Thus, the emphasis on fractions has been shifted to better reflect classroom instruction.

C10. Division of Fractions

This scale tests the pupil's ability to divide a proper fraction or whole number by a proper fraction.

Sample:

$8 \div \frac{4}{5} = \square$	A. $\frac{1}{6}$ C. 10 B. $9\frac{4}{5}$ D. 11

Pilot Data:

Number of Items	8
Number of Students	79
Mean Total Score	1.05
Standard Deviation	1.14
Mean Proportion Correct	.14
Internal Consistency	.36

Ability-OTL Matrix:

MEAN / COUNT	RAVSTRAT	1	4	ROW TOTAL
1	LOW STRATA	1.1 22	0.0 1	1.0 23
2	MIDDLE STRATA	1.6 21	0.0 0	1.6 21
3	HIGH STRATA	0.7 23	0.4 5	0.6 28
	COLUMN TOTAL	1.1 66	0.3 6	1.0 72

Revisions: This scale has been dropped. See comments on C9.

C11. Addition and Subtraction of Decimals

This scale tests the pupil's ability to add and subtract decimals involving tenths, hundreths, and thousandths. All items are presented in the vertical format.

Sample:

8.36 -5.61	A. .270 C. .335 B. 2.75 D. 3.75

Pilot Data:

Number of items	10
Number of Students	79
Mean Total Score	7.84
Standard Deviation	2.65
Mean Proportion Correct	.78
Internal Consistency	.85

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	4	5	ROW TOTAL
RAVSTRAT 1 LOW STRATA	4.5 4	9.3 4	4.8 6	4.5 4	7.4 5	6.0 23
2 MIDDLE STRATA	6.8 6	10.0 2	8.0 10	9.7 3	0.0 0	8.1 21
3 HIGH STRATA	6.3 3	9.7 3	7.9 10	9.0 9	9.7 3	8.5 28
COLUMN TOTAL	6.0 13	9.6 9	7.2 26	8.0 16	8.3 8	7.6 72

Revisions: Format has been changed to include both the horizontal and vertical form plus the use of the $ sign. The thousandths place has been dropped as inappropriate for grade 5. Five items from this pilot scale have been kept with the remainder being modified or dropped. New sample Item:

$28.54 - $5.64 = ☐

A. $22.90 B. $22.10 C. $2.29 $ 12.90

C12. Number sentences

This scale tests the pupil's ability to solve number sentences involving equalitites and inequalities.

Sample:

What symbol goes in the ◯ ?

$7 + 8 \bigcirc 45 \times \frac{1}{3}$

A. $<$ B. $>$ C. $=$ D. $+$

Pilot Data:

Number of Items	10
Number of Students	79
Mean Total Score	4.41
Standard Deviation	2.43
Mean Proportion Correct	.44
Internal Consistency	.74

Ability-OTL Matrix:

RAWSTRAT	MEAN COUNT	1	3	ROW TOTAL
1 LOW STRATA		3.8 17	3.2 6	3.7 23
2 MIDDLE STRATA		4.0 11	5.0 10	4.5 21
3 HIGH STRATA		4.7 18	6.2 10	5.3 28
COLUMN TOTAL		4.2 46	5.0 26	4.5 72

Revisions: This scale has been dropped. Teachers questioned the purpose of testing this objective and also indicated that this was more likely to be taught at the earlier grade levels. Time considerations were also a factor in dropping this scale.

C13. Speed Test - Computation of Whole Numbers (2 minutes)

This scale tests the pupil's ability to recall basic facts and perform simple computational problems rapidly. There are 3 items each for addition and subtraction and 4 items each for multiplication and division. Five items require regrouping.

Sample:

43 x 3	A. 169	C. 129
	B. 126	D. 40

Pilot Data:

Number of Items	14
Number of Students	79
Mean Total Score	8.97
Standard Deviation	3.17
Mean Proportion Correct	.64
Internal Consistency	.84

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	4	ROW TOTAL
1 LOW STRATA	8.1 8	7.7 6	8.0 9	8.0 23
2 MIDDLE STRATA	12.3 3	8.8 10	7.1 8	8.7 21
3 HIGH STRATA	9.3 7	9.9 10	10.7 11	10.1 28
COLUMN TOTAL	9.3 18	9.0 26	8.8 28	9.0 72

Revisions: None

N1. Decimal Place Value

This scale tests the pupil's knowledge of decimal place value involving tenths, hundredths, and thousandths. Two items involve the recognition of an equivalent decimal fraction.

Sample:

$\frac{6}{10}$ is equivalent to which decimal fraction?

A. 6.00
B. .0006
C. .6
D. .06

Pilot Data:

Number of Items	6
Number of Students	79
Mean Total Score	1.3
Standard Deviation	1.3
Mean Proportion Correct	.22
Internal Consistency	.52

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	3	4	ROW TOTAL
1 LOW STRATA	1.0 4	1.2 6	1.6 8	0.8 5	1.2 23
2 MIDDLE STRATA	1.0 6	0.9 10	1.4 5	0.0 0	1.0 21
3 HIGH STRATA	1.3 3	0.9 10	2.6 12	1.3 3	1.7 28
COLUMN TOTAL	1.1 13	1.0 26	2.0 25	1.0 8	1.4 72

Revisions: Scale expanded to 10 items and includes whole number place value. Some pilot items revised.

N2. Properties

This scale tests the pupil's knowledge of the commutative, associative and distributive properties. Also includes items on identity elements. All items are open ended. Seven items involved frations. This data was hand scored and the standard deviation and internal consistency figures are not available.

Sample:

$$\frac{5}{27} \times \frac{\square}{\triangle} = \frac{47}{51} \times \frac{5}{27}$$

$$\frac{\square}{\triangle} = \text{—}$$

Pilot Data:

Number of Items	10
Number of Students	79
Mean Total Score	3.4
Mean Proportion Correct	.34

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	3	ROW TOTAL
1 LOW STRATA	1.8 17	4.7 6	2.6 23
2 MIDDLE STRATA	3.2 11	4.9 10	3.9 21
3 HIGH STRATA	3.2 18	6.2 10	4.3 28
COLUMN TOTAL	2.7 46	4.6 26	3.4 72

Revisions: Fractional format was dropped.

N3. Prime Numbers

Tests pupil's ability to identify a prime number.

Sample:

> Which of these is a prime number?
>
> A. 8 C. 25
>
> B. 6 D. 19

Pilot Data:

Number of Items	3
Number of Students	79
Mean Total Score	.53
Standard Deviation	.74
Mean Proportion Correct	.18
Internal Consistency	.34

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
RAWSTRAT				
1 LOW STRATA	0.6 / 17	0.7 / 6	0.0 / 0	0.6 / 23
2 MIDDLE STRATA	0.4 / 11	0.8 / 8	0.5 / 2	0.5 / 21
3 HIGH STRATA	0.2 / 18	0.9 / 7	2.0 / 3	0.6 / 28
COLUMN TOTAL	0.4 / 46	0.8 / 21	1.4 / 5	0.6 / 72

Revisions: This scale has been dropped as teacher feedback and data results indicate that it does not reflect classroom instruction.

N4(a). Multiples

Tests pupil's ability to identify the multiples of a number

Sample:

> Which set below is a set of multiples of 6?
>
> A. {1, 2, 3, 4, 5, 6}
> B. {0, 3, 6, 9, 12}
> C. {1, 2, 3, 6}
> D. {0, 6, 18, 24}

Pilot Data:

Number of Items	3
Number of Students	79
Mean Total Score	.84
Standard Deviation	.96
Mean Proportion Correct	.28
Internal Consistency	.52

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	0.8 / 12	0.5 / 5	0.0 / 5	0.6 / 23
2 MIDDLE STRATA	1.0 / 5	1.8 / 10	0.2 / 6	1.1 / 21
3 HIGH STRATA	1.2 / 10	1.1 / 10	0.3 / 8	0.9 / 28
COLUMN TOTAL	1.0 / 27	1.2 / 26	0.2 / 19	0.9 / 72

Revisions: This scale has been dropped along with C17 and C18 which are extensions of this scale. Teacher and consultant feedback indicate that very little instructional time is given to the teaching of multiples.

N4(b). Common multiples

This scale is an extension of C16 and tests the pupil's knowledge of common multiples.

Sample:

> Which number is a multiple of both 8 and 12?
>
> A. 36 C. 16
>
> B. 48 D. 12

Pilot Data:

Number of Items	3
Number of Students	79
Mean Total Score	.65
Standard Deviation	.68
Mean Proportion Correct	.22
Internal Consistency	- .02

Ability-OTL Matrix:

MEAN / COUNT	1	2	3	ROW TOTAL
RAVSTRAT				
1 LOW STRATA	2.7 / 12	0.5 / 6	1.0 / 5	1.7 / 23
2 MIDDLE STRATA	0.6 / 5	0.5 / 10	0.7 / 6	0.6 / 21
3 HIGH STRATA	0.6 / 10	0.7 / 10	0.9 / 8	0.7 / 28
COLUMN TOTAL	1.5 / 27	0.6 / 26	0.8 / 19	1.0 / 72

Revisions: This scale has been dropped. See comments for C16.

N4(c). Least Common Multiples

This scale is an extension of C16 and C17 and tests the pupil's knowledge of least common multiples.

Sample:

> What is the least common multiple of 6 and 6?
>
> A. 18 B. 36 C. 9 D. 54

Pilot Data:

Number of Items	3
Number of Students	79
Mean Total Score	1.05
Standard Deviation	.98
Mean Proportion Correct	.35
Internal Consistency	.54

Ability-OTL Matrix:

MEAN / COUNT	1	2	ROW TOTAL
RAVSTRAT			
1 LOW STRATA	3.1 / 12	0.5 / 11	1.9 / 23
2 MIDDLE STRATA	1.2 / 5	1.3 / 16	1.3 / 21
3 HIGH STRATA	0.7 / 10	1.3 / 18	1.1 / 28
COLUMN TOTAL	1.9 / 27	1.1 / 45	1.4 / 72

Revisions: Scale has been dropped. See comments for C16.

N5(a). Factors

This scale is designed to test the pupil's knowledge of the factors of a number.

Sample:

> If 8 is one factor of 32, what is the other factor?
>
> A. 1 B. 3 C. 4 D. 6

Pilot Data:

Number of Items	3
Number of Students	79
Mean Total Score	1.05
Standard Deviation	.98
Mean Proportion Correct	.35
Internal Consistency	.45

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	1.4 12	1.1 10	1.0 1	1.3 23
2 MIDDLE STRATA	1.0 5	0.9 14	1.5 2	1.0 21
3 HIGH STRATA	1.4 10	1.2 10	0.9 8	1.2 28
COLUMN TOTAL	1.3 27	1.0 34	1.0 11	1.1 72

Revisions: This scale, along with C20 and C21, has been dropped. It appears that very little instruction is given to number theory , i.e.: prime numbers, multiples, and factors. A new scale of 10 items has been developed on factors and prime factorization which is more graduated and hopefully will be more sensitive to the students' knowledge of factoring. This is the only scale on number theory in the revised battery.

N5(b). Common Factors.

This is an extension of C19 and involves the identification of the common factors of a set of numbers.

Sample:

> Which set names the common factors of 28 and 35?
>
> A. {1, 7} C. {1, 4}
>
> B. {1, 5} D. {1, 4, 5, 7}

Pilot Data:

Number of Items	3
Number of Students	79
Mean Total Score	.68
Standard Deviation	.80
Mean Proportion Correct	.23
Internal Consistency	.34

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	0.7 12	0.7 11	0.0 0	0.7 23
2 MIDDLE STRATA	0.8 5	0.8 14	1.5 2	0.9 21
3 HIGH STRATA	0.9 10	0.4 15	1.3 3	0.7 28
COLUMN TOTAL	0.8 27	0.6 40	1.4 5	0.7 72

Revisions: Scale has been dropped. See C19 for comments.

N5(c). Greatest Common Factor

This scale is an extension of C19 and C20 and involves the identification of the greatest common factor of a set of numbers.

Sample:

> What is the greatest common factor of 14 and 21?
>
> A. 7 B. 2 C. 3 D. 14

Pilot Data:

Number of Items	4
Number of Students	79
Total mean score	.89
Standard Deviation	1.04
Mean Proportion Correct	.22
Internal Consistency	.52

Ability-OTL Matrix:

MEAN COUNT / RAVSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	1.7 12	0.7 6	0.0 5	1.0 23
2 MIDDLE STRATA	1.6 5	1.4 8	0.3 8	1.0 21
3 HIGH STRATA	0.9 10	1.1 7	0.4 11	0.8 28
COLUMN TOTAL	1.4 27	1.1 21	0.3 24	0.9 72

Revisions: Scale has been dropped. See C19.

N6. Prime Factorization

This scale is somewhat of an extension of scales C19 to C21. It tests the pupil's ability to recognize or complete the prime factorization of a number.

Sample:

> Which number names the missing prime factor?
>
> 42 = 2 x ☐ x 7
>
> A. 3 B. 7 C. 4 D. 2

Pilot Data:

Number of items	4
Number of students	79
Mean Total Score	.59
Standard Deviation	.85
Mean Proportion Correct	.15
Internal Consistency	.45

Ability-OTL Matrix:

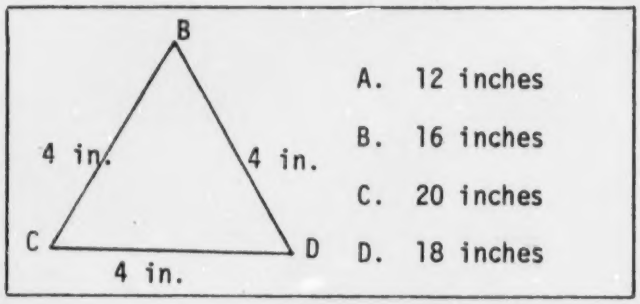
MEAN / COUNT	RAVSTRAT	1	2	3	ROW TOTAL
1	LOW STRATA	0.4 17	1.5 6	0.0 0	0.7 23
2	MIDDLE STRATA	0.4 11	0.8 8	1.0 2	0.6 21
3	HIGH STRATA	0.3 18	0.9 7	1.3 3	0.5 28
	COLUMN TOTAL	0.3 46	1.0 21	1.2 5	0.6 72

Revisions: Two items kept and included in revised scale (see comments for N5(a)).

G1. Perimeter

This scale tests the pupil's ability to compute the perimeter of a polygon using standard or arbitrary units of measure (3 items each).

Sample:



Pilot Data:

Number of Items	6
Number of Students	78
Mean Total Score	2.97
Standard Deviation	1.15
Mean Proportion Correct	.50
Internal Consistency	.34

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	2.2 5	3.5 8	1.8 6	2.6 19
2 MIDDLE STRATA	2.8 9	3.4 7	3.3 7	3.1 23
3 HIGH STRATA	3.3 6	3.4 8	3.2 12	3.3 26
COLUMN TOTAL	2.8 20	3.4 23	2.9 25	3.0 68

Revisions: Improved graphics and defined arbitrary units within item question. Added one item from the ETS Grade 5 test battery and dropped one pilot item. Combined with other metric geometry objectives to form one scale.

G2 & 4. Area

This scale test's the pupil's ability to find the area of a polygon using standard or arbitrary units (5 items each) and to compute the surface area of a rectangular prism. These "subtests" were treated as one scale for the Pilot Data but the OTL Matrices are given separately for area of polygons (Sample A) and surface area of rectangular prisms (Sample B).

Pilot Data:

Number of Items	14
Number of Students	78
Mean Total Score	4.58
Standard Deviation	1.93
Mean Proportion Correct	.33
Internal Consistency	.44

Sample (G2):

What is the area of the rectangular region ABCD?

A. 4 sq. units
C. 8 sq. units
B. 6 sq. units
D. 12 sq. units

Number of Items	10
Mean Proportion Correct	.42

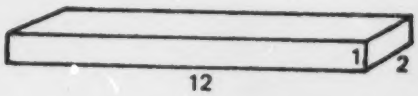
Ability-OTL Matrix:

MEAN COUNT	1	2	ROW TOTAL
RAWSTRAT			
1 LOW STRATA	3.5 2	3.9 17	3.8 19
2 MIDDLE STRATA	3.6 5	4.5 18	4.3 23
3 HIGH STRATA	4.0 2	4.4 24	4.3 26
COLUMN TOTAL	3.7 9	4.3 59	4.2 68

G2 & 4 Cont'd.

Sample (G4):

What is the surface area of this rectangular prism?



A. 76 sq. units C. 52 sq. units

B. 48 sq. units D. 38 sq. units

Number of items 4
Mean Proportion Correct .09

Ability-OTL Matrix:

MEAN COUNT / RASTRAT	1	2	ROW TOTAL
1 LOW STRATA	0.3 18	1.0 1	0.3 19
2 MIDDLE STRATA	0.5 21	0.5 2	0.5 23
3 HIGH STRATA	0.4 24	0.5 2	0.4 26
COLUMN TOTAL	0.4 63	0.6 5	0.4 68

Revisions: Subtest on the surface area of rectangular prisims has been dropped as data results and teacher feedback indicate that it is inappropriate for Grade 5. Reduced number of items to 3 on recognizing area and 3 on computing area. Defined the arbitrary unit, combined with the other metric geometry objectives to form one scale.

G3 & 5. Volume

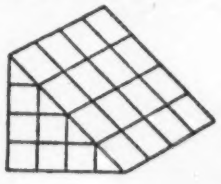
This scale tests the pupil's ability to find or compute the volume of a space figure using arbitrary (6 items) or standard units of measure (4 items). These subtests have been combined for the Pilot Data but separate OTL-Matrices are given for each.

Pilot Data:

Number of Items	10
Number of Students	78
Mean Total Score	2.88
Standard Deviation	2.01
Mean Proportion Correct	.29
Internal Consistency	.61

Sample - Arbitrary Units (G3):

How many cubic units are in this space figure



A. 12 B. 16 C. 32 D. 64

Number of Items	5
Mean Proportion Correct	.36

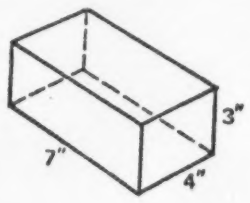
Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	ROW TOTAL
1 LOW STRATA	2.7 13	2.1 7	2.5 19
2 MIDDLE STRATA	2.0 14	2.4 9	2.2 23
3 HIGH STRATA	1.3 12	1.9 14	1.6 26
COLUMN TOTAL	2.0 38	2.1 30	2.0 68

G3 & 5 Cont'd.

Sample G5 - Standard Units:

What is the volume of a shoe box which measures 4 inches by 7 inches by 3 inches?



A. 56 cu. inches C. 25 cu. inches

B. 84 cu. inches D. 33 cu. inches

Number of Items 4
Mean Proportion correct .19

Ability-OTL Matrix:

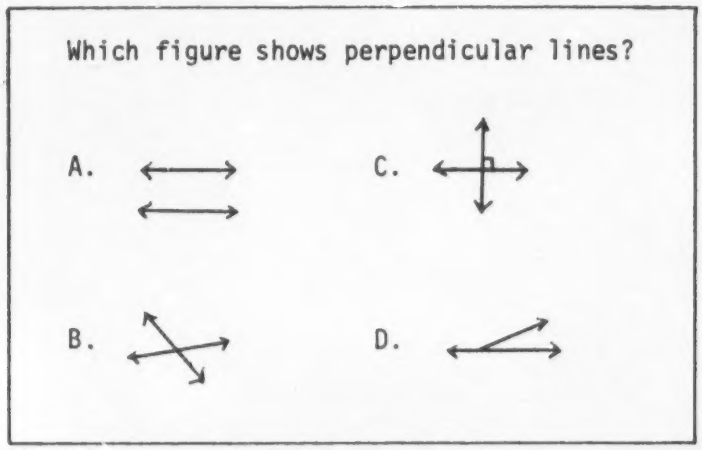
MEAN COUNT / MAINSTRAT	1	ROW TOTAL
1 LOW STRATA	0.6 19	0.6 19
2 MIDDLE STRATA	0.6 23	0.6 23
3 HIGH STRATA	0.9 26	0.9 26
COLUMN TOTAL	0.7 68	0.7 68

Revisions: Reduced scale to 5 items. Three items on "recognizing" volume and 2 items on computing volume. Defined the arbitrary unit. Combined with other metric geometry subtests to form one scale.

G6. Lines

This scale tests the pupil's knowledge of parallel lines, line segments, lines and angles.

Sample:



Pilot Data:

Number of Items	10
Number of Students	78
Mean Total Score	4.45
Standard Deviation	2.16
Mean Proportion Correct	.45
Internal Consistency	.60

OTL-Ability Matrix:

MEAN COUNT / RAWSTRAT	1	2	ROW TOTAL
1 LOW STRATA	3.8 10	5.3 9	4.5 19
2 MIDDLE STRATA	4.5 14	5.1 9	4.7 23
3 HIGH STRATA	4.3 16	4.5 10	4.4 26
COLUMN TOTAL	4.3 40	5.0 28	4.5 68

Revisions: Reduced scale to 8 items and combined it with 4 items on plane geometric figures to create one non-metric geometry scale.

G7. Geometric Figures

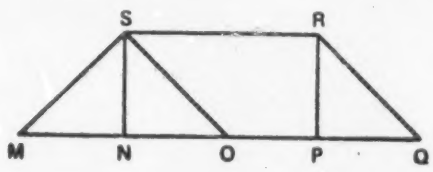
This scale tests the pupil's knowledge of plane and solid geometric figures. There are 10 items on plane figures (G7-a) and 6 items on solid figures (G7-b). OTL-Matrices are given for each of these subtests and Pilot Data is reported on the scale as a whole.

Pilot Data:

Number of Items	16
Number of Students	78
Mean Total Score	10.65
Standard Deviation	3.03
Mean Proportion Correct	.67
Internal Consistency	.73

Sample (G7-a):

In the figure below, which of the following names a rectangle?



A. SRPO

B. MSO

C. SRPN

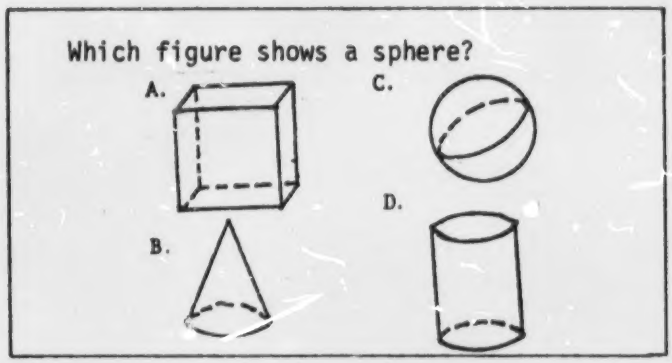
D. SRQO

Ability-OTL Matrix:

MEAN COUNT / RASTRAT	1	2	ROW TOTAL
LOW STRATA 1	5.2 / 5	6.3 / 14	6.0 / 19
MIDDLE STRATA 2	7.2 / 5	6.5 / 14	6.8 / 23
HIGH STRATA 3	6.7 / 6	6.1 / 20	6.3 / 26
COLUMN TOTAL	6.5 / 20	6.3 / 48	6.4 / 68

G7 Cont'd.

Sample (G7-b):



Ability-OTL Matrix:

RASTRAT	MEAN CCUNT	1	2	ROW TOTAL
1 LOW STRATA	3.5	11	8	3.7 19
2 MIDDLE STRATA	4.7	16	7	4.3 23
3 HIGH STRATA	4.3	18	8	4.4 26
CCLUMN TOTAL	4.2	45	23	4.2 68

Revisions: Four items were taken from the G7-a, plane figures, and combined with G6 to form one non-metric geometry scale. The remaining items were dropped along with the items on solid geometric figures. Many of the items on plane figures reflected instruction given prior to Grade 5 and few teachers covered solid geometric figures.

G8. Circle

This scale tests the pupil's knowledte of the parts of a circle: radius, diameter, arc and chord. Items testing the quantative relationship between the radius and diameter are also included.

Sample:

Which line segment names a radius?

A) CB

B) $\overline{CD}$

C) AC

D) $\overline{OA}$

Pilot Data:

Number of Items	10
Number of Students	78
Mean Total Score	2.81
Standard Deviation	1.91
Mean Proportion Correct	.28
Internal Consistency	.51

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT		1	ROW TOTAL
LOW STRATA	1	3.2 19	3.2 19
MIDDLE STRATA	2	3.3 23	3.3 23
HIGH STRATA	3	3.2 26	3.2 26
COLUMN TOTAL		3.2 68	3.2 68

Revisions: This scale has been dropped as not being reflective of classroom instruction.

S1 & 2. Number pairs/Line and Bar Graphs

This scale tests the pupil's ability to read and interpret line and bar graphs and to identify number pairs on a number plane. The pilot data combines S1 and 2 and OTL-Matrices are given for each scale.

Pilot Data:

Number of Items	10
Number of Students	78
Mean Total Score	5.42
Standard Deviation	3.02
Mean Proportion Correct	.54
Internal Consistency	.82

Sample A: S1

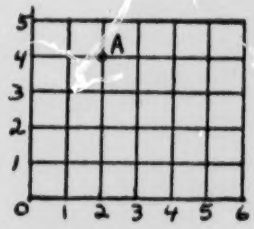
What is the number pair for point A?

A. (3, 4)

B. (2, 4)

C. (4, 2)

D. (4, 1)



Number of Items	4
Mean Proportion Correct	.53

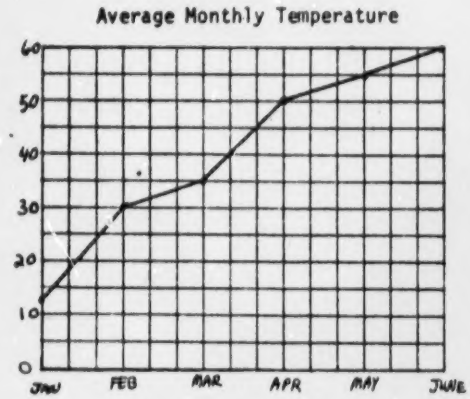
Ability-OTL Matrix:

MEAN SCORE / ROW TOTAL	1	3	ROW TOTAL
RATSTRAT			
1 LOW STRATA	1.2 11	2.6 8	1.8 19
2 MIDDLE STRATA	1.4 16	3.4 7	2.0 23
3 HIGH STRATA	1.7 18	3.5 8	2.3 26
COLUMN TOTAL	1.5 45	3.2 23	2.1 68

S1 & 2 Cont'd

Sample B: S2

Using this graph, find the average temperature for the month of May.



A. 62°

B. 50°

C. 55°

D. None of these

Number of Items 6
Mean Proportion Correct .56

Ability-OTL Matrix:

MEAN COUNT / ABILITY STRATA	1	2	3	ROW TOTAL
1 LOW STRATA	2.5 11	0.0 0	2.9 8	2.7 19
2 MIDDLE STRATA	3.2 13	3.7 3	4.3 7	3.6 23
3 HIGH STRATA	3.1 13	2.6 5	4.8 8	3.5 26
COLUMN TOTAL	3.0 37	3.0 8	4.0 23	3.3 68

Revisions: Two graphs (5 items modified to increase their level of difficulty. One item from the Math ETU and one from the ETS test battery have been included.

S3. Average, Range, Median

This scale tests the pupil's ability to find the average, range, or median of a set of numbers.

Sample:

> What is the average of the numbers 2, 6, 12, 24 and 28?
>
> A. 15 B 26 C. 90 D. 19

Pilot Data:

Number of Items	6
Number of Students	78
Mean Total Score	1.24
Standard Deviation	1.24
Mean Proportion Correct	.21
Internal Consistency	.44

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	ROW TOTAL
1 LOW STRATA	1.3 18	0.0 1	1.3 19
2 MIDDLE STRATA	1.2 18	2.8 5	1.5 23
3 HIGH STRATA	1.1 19	0.9 7	1.0 26
COLUMN TOTAL	1.2 55	1.5 13	1.3 68

Revisions: This scale has been dropped. Range and median are not usually covered in 5th grade. Items involving averages are already contained in one of the word problem scales and will be retained which should sufficiently test this objective.

S4. Number Patterns

This scale tests the pupil's ability to determine the numerical relationship with a sequence of letter or numbers. Pupil must choose the number (letter) which is missing and/or state the rule by which it can be found.

Sample:

> Following the pattern in the sequence below, what letter goes in the [ ] ?
>
> A, X, B, X, C, [ ]
>
> A) D B) X C) A D) Z

Pilot Data:

Number of Items	6
Number of Students	78
Mean Total Score	4.28
Standard Deviation	1.64
Mean Proportion Correct	.71
Internal Consistency	.61

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	3	ROW TOTAL
1 LOW STRATA	3.2 6	3.5 12	4.0 1	3.4 19
2 MIDDLE STRATA	4.1 7	4.4 14	6.0 2	4.4 23
3 HIGH STRATA	4.5 12	4.3 12	6.0 2	4.5 26
COLUMN TOTAL	4.1 25	4.1 38	5.6 5	4.2 68

Revisions: Expanded to 10 items with increased difficulty. Dropped 3 items which were more appropriate for Grade 3-4.

A1. One step Word Problems

This scale tests pupil's ability to solve one step word problems involving any basic process.

Sample:

> John has 96 marbles which he wants to put into sets of 8. How many complete sets can he make?
>
> A. 10 B. 8 C. 12 D. 88

Pilot Data:

Number of items	5
Number of students	78
Mean Total Score	2.04
Standard Deviation	1.21
Mean Proportion Correct	.41
Internal Consistency	.36

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	2	4	5	ROW TOTAL
1 LOW STRATA	1.6 14	2.0 4	2.0 1	1.7 19
2 MIDDLE STRATA	2.1 14	1.7 7	2.5 2	2.0 23
3 HIGH STRATA	2.5 20	2.0 4	3.0 2	2.5 26
COLUMN TOTAL	2.1 48	1.9 15	2.6 5	2.1 68

Revisions: Replaced one item which involved fractions.

A2. Two Step Word Problems

This scale is an extension of A1 and tests the students ability to solve difficult word problems involving two computational steps.

Sample:

> Anne has 12 dimes. George has 5 dimes less than Anne. Which equation can be used to tell how many dimes they have altogether?
>
> A. 12 - 5 = ☐ C. 12 + (12-5) = ☐
>
> B. 12 + 5 = ☐ D. (12+5) + 12 = ☐

Pilot Data:

Number of Items	9
Number of Students	78
Mean Total Score	5.31
Standard Deviation	2.64
Mean Proportion Correct	.59
Internal Consistency	.60

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	5	ROW TOTAL
1 LOW STRATA	1.9 14	0.0 0	1.8 5	1.9 19
2 MIDDLE STRATA	3.2 11	4.3 3	3.0 9	3.3 23
3 HIGH STRATA	3.3 15	6.4 5	4.2 6	4.1 26
COLUMN TOTAL	2.8 40	5.6 8	3.0 20	3.2 68

Revisions: Two items, involving rate, distrance, and time, were dropped to decrease the emphasis on problems involving rate. One new item was added.

A3. Word Problems Involving Metric Geometry

This scale tests the pupil's ability to solve word problems involving the computation of perimeter, area, and volume (1 item each).

Sample:

> Our neighborhood park has a rectangular shape and is 114 feet wide and 142 feet long. If you walked all the way around the park, how far would you have walked?
>
> A. 3,224 feet C. 512 feet
>
> B. 621 feet D. 284 feet

Pilot Data:

Number of Items	3
Number of Students	78
Total Mean Score	.71
Standard Deviation	.83
Mean Proportion Correct	.24
Internal Consistency	-.02

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	1	2	ROW TOTAL
1 LOW STRATA	0.5 11	0.4 8	0.4 19
2 MIDDLE STRATA	0.3 12	0.7 10	0.5 23
3 HIGH STRATA	0.6 13	0.5 13	0.6 26
COLUMN TOTAL	0.5 37	0.5 31	0.5 68

Revisions: This scale has been dropped due to time considerations and that these concepts will be sufficiently tested with scale G1 (a, b, and c).

A4. Word Problems Involving Common Measures

This scale tests the pupil's ability to solve word problems involving common linear and liquid measures, time and money.

Sample:

> During a track meet, Bill and Dave participated in the broad jump event. Bill jumped 1 yard 2 feet 3 inches and Dave jumped 2 yards 2 inches. How much further did Dave Jump than Bill?
>
> A. 9 inches C. 1 foot
>
> B. 1 foot 1 inch D. 11 inches

Pilot Data:

Number of Items	10
Number of Students	78
Mean Total Score	3.26
Standard Deviation	1.97
Mean Proportion Correct	.33
Internal Consistency	.54

Ability-OTL Matrix:

MEAN COUNT / ABILITY STRATA	OTL 2	3	ROW TOTAL
1 LOW STRATA	2.6 14	3.2 5	2.7 19
2 MIDDLE STRATA	3.9 14	3.0 9	3.5 23
3 HIGH STRATA	3.7 20	3.7 6	3.7 26
COLUMN TOTAL	3.4 48	3.3 20	3.4 68

Revisions: Three items dropped due to time considerations. One item modified to eliminate fractional format.

A5. Word Problems Involving Averages

This scale tests the pupil's ability to solve word problems which involve the use of an average or require the computing of an average.

Sample:

> Last Friday Jeff bowled games scoring 171, 149, 206, and 210. What was his average score?
>
> A. 136 B. 184 C. 673 D. 368

Pilot Data:

Number of Items	3
Number of Students	78
Mean Total Score	.77
Standard Deviation	.85
Mean Proportion Correct	.26
Internal Consistency	.33

Ability-OTL Matrix:

MEAN COUNT / ROW TOTAL	1	2	ROW TOTAL
RA\STRAT			
1 LOW STRATA	0.4 19	0.0 0	0.4 19
2 MIDDLE STRATA	0.9 20	1.3 3	1.0 23
3 HIGH STRATA	0.9 21	1.6 5	1.0 26
COLUMN TOTAL	0.7 60	1.5 8	0.8 68

Revisions: All items incorporated into the two-step word problem scale (A2).

A6. Computation/Word Problem Cross-reference

This scale tests the pupil's ability to perform the computational process or part of that process required by the cross-referenced word problem. Pilot Data is on the total scale. OTL matrices were done for the multiplication and division problems only.

Pilot Data:

Number of Items	14
Number of Students	78
Mean Total Score	7.28
Standard Deviation	3.32
Mean Proportion Correct	.52
Internal Consistency	.79

Sample A (Multiplication):

48 x24	A. 1056	B. 1152
	C. 1052	D. 1504

Number of Items	4
Mean Proportion Correct	.60

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	2	3	4	5	ROW TOTAL
1 LOW STRATA	4.0 1	2.0 2	1.6 8	1.6 8	1.8 19
2 MIDDLE STRATA	3.5 2	2.0 2	2.0 7	2.6 12	2.4 23
3 HIGH STRATA	3.5 2	3.0 2	2.4 8	2.9 14	2.8 26
COLUMN TOTAL	3.6 5	2.3 6	2.0 23	2.5 34	2.4 68

A6. (continued)

Sample B (Division (corresponds to A1 sample item)):

$8\overline{)96}$

A. 12 B. 8 C. 10 D. 88

Number of Items 5
Mean Proportion Correct .53

Ability-OTL Matrix:

MEAN COUNT / RAWSTRAT	3	4	5	ROW TOTAL
1 LOW STRATA	2.0 5	1.0 4	1.3 6	1.6 19
2 MIDDLE STRATA	2.4 5	1.4 7	3.4 7	2.4 23
3 HIGH STRATA	3.0 10	3.5 4	4.2 12	3.6 26
COLUMN TOTAL	2.5 28	1.9 15	3.3 25	2.6 68

Revisions: Deleted items which corresponded to deleted word problems and added replacement items.

APPENDIX B

Subtest Intercorrelations

Each table in this appendix contains intercorrelations between the subtests given to one group of students. Subtests are identified according to the code numbers in Tables I through IV. The correlations are not corrected for attenuation. Internal consistency estimates of reliability are indicated in the diagonal entries. Where reliability estimates are not present, a subset of items from a subtest have been presented as a separate scale. This happened when Opportunity to Learn was analyzed separately for subsets of items. In many cases only a small number of items are included in these subsets. See the subtest summary pages for further details about scale size and reliability.

Selected Scale Intercorrelations - Grade 2 Reading

GROUP A
N=67

		D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	C1	C2	C3	C4	M1	M2	M3	M4	R	V
Sing. Cons.	D1	(.95)																				
Db. Cons.I	D2	.38	(.77)																			
Db. Cons. - F	D3	.68	.38	(.80)																		
Long-e	D4	.24	.70	.33	(.73)																	
Digraph	D5	.44	.69	.37	.74	(.81)																
Diphthong	D6	.35	.67	.51	.76	.70	(.80)															
Vowel with r	D7	.58	.57	.63	.62	.69	.80	(.79)														
Internal	D8	.25	.44	.49	.53	.47	.74	.67	(.78)													
Cons. Sub.	D9	.16	.55	.36	.68	.51	.76	.52	.63	(.83)												
Silent	D10	.10	.42	.34	.61	.37	.62	.53	.64	.52	(.88)											
Var. Cons.	D11	.00	.24	.24	.37	.24	.52	.43	.51	.49	.42	(.65)										
Context - Cons.	C1	.33	.43	.53	.60	.58	.79	.78	.76	.54	.57	.44	(.90)									
Context - Syntax	C2	.14	.41	.34	.47	.45	.53	.36	.38	.48	.37	.25	.52	(.85)								
Context - Wd. Struc.	C3	.27	.34	.45	.55	.52	.69	.62	.57	.55	.51	.47	.63	.56	(.75)							
Context - Pronoun	C4	.08	.30	.30	.47	.38	.60	.52	.57	.51	.63	.46	.60	.46	.54	(.73)						
ID Root	M1	.25	.43	.48	.47	.54	.64	.58	.50	.52	.45	.49	.44	.45	.56	.50	(.89)					
Compunds	M2	.64	.63	.58	.55	.66	.73	.72	.55	.56	.40	.26	.59	.63	.58	.40	.59	(.94)				
Divide Root	M3	.44	.58	.60	.64	.70	.86	.81	.70	.64	.62	.47	.74	.55	.68	.61	.76	.75	(.88)			
No. Syl	M4	.22	.33	.30	.52	.54	.64	.59	.47	.45	.59	.36	.61	.43	.64	.63	.45	.52	.59	(.85)		
Ravens	R	.03	.10	.10	.16	-.01	.20	.21	.13	.24	.39	.16	.13	.06	.33	.21	.23	.05	.34	.08	-	
Vocab.	V	.23	.51	.46	.60	.61	.72	.65	.67	.66	.70	.44	.72	.54	.71	.60	.60	.57	.75	.57	.4C	(.89)

Selected Scale Intercorrelations - Grade 2 Reading
GROUP B (N=72)

		M1	M2	M3	W5	T1	T2	T3	T4	TG	A	R	V
Multi-Mng	M1	(.78)											
Unfamiliar	M2	.60	-										
Pron. Ref.	M3	.48	.42	(.58)									
Contractions	W5	.53	.52	.32	(.81)								
Recall	T1	.71	.68	.65	.45	(.73)							
Translation	T2	.61	.54	.56	.32	.74	(.77)						
Integration	T3	.59	.56	.58	.44	.75	.70	(.70)					
Extrapolation	T4	.66	.66	.56	.44	.80	.76	.72	(.68)				
Main Idea	TG	.58	.55	.49	.38	.78	.62	.70	.81	-			
Applied	A	.75	.72	.60	.50	.81	.76	.70	.77	.63	(.89)		
Ravens	R	.47	.40	.37	.34	.41	.34	.29	.35	.28	.51	-	
Vocab.	V	.69	.67	.61	.50	.72	.64	.60	.73	.65	.79	.43	(.89)

Selected Scale Intercorrelations - Grade 5 Reading
GROUP A
N=73

		D1	D2	D3	D4	C1	C2	C3	W1	W2	W3	W4	W5	A2	A3	A5	A6	R1	R2	R3	M1	R	V
Var. Cons.	D1	(.79)																					
L/S Vowel	D2	.58	(.83)																				
Vowel + r	D3	.65	.56	(.87)																			
Silent letter	D4	.55	.41	.47	(.93)																		
CC-Meaning	C1	.73	.58	.63	.50	(.69)																	
CC-Form	C2	.69	.51	.68	.46	.67	(.77)																
CC-Pronoun	C3	.58	.54	.59	.40	.61	.64	(.78)															
ID-root	W1	.61	.67	.51	.46	.58	.50	.58	(.78)														
ID-parts	W2	.65	.45	.52	.42	.55	.48	.52	.51	(.84)													
First Syl.	W3	.67	.71	.61	.49	.63	.60	.61	.56	.55	(.82)												
Last Syl.	W4	.48	.44	.33	.32	.29	.42	.35	.57	.27	.53	(.73)											
No. Syl.	W5	.63	.61	.59	.40.	.60	.61	.43	.65	.43	.64	.61	(.82)										
Signs	A2	.55	.46	.49	.39	.51	.45	.47	.39	.38.	.51	.35	.31	-									
Directions	A3	.55	.44	.47	.34	.61	.38	.41	.59	.46	.46	.44	.58	.34	-								
Map	A5	.47	.47	.34	.26	.52.	.48	.43	.54	.39	.44	.42	.55	.27	.47	-							
Graph	A6	.51	.57	.38	.28	.57	.56	.37	.43	.36	.60	.47	.57	.39	.50	.49	-						
Ref. In.	R1	.60	.60	.52	.44	.59	.50	.42	.47	.47	.52	.34	.54	.53	.44	.40	.48	-					
Ref. Betwn.	R2	.52	.53	.55	.34	.50	.54	.37	.45	.44	.47	.28	.51	.43	.37	.34	.47	.56	-				
Outline	R3	.46	.39	.39	.28	.40	.41	.25	.33	.32	.44	.22	.31	.18	.35	.28	.29	.34	.38	-			
Multi-mng	M1	.61	.61	.76	.56	.63	.66	.56	.57	.49	.64	.41	.55	.48	.49	.47	.48	.63	.60	.48	-		
Ravens	R	.44	.53	.43	.32	.30	.46	.32	.30	.43	.49	.28	.30	.45	.18	.33	.46	.48	.42	.25	.43	-	
Vocab.	V	.73	.67	.73	.59	.67	.70	.60	.65	.57	.68	.45	.54	.56	.46	.46	.48	.69	.62	.53	.79	.52	.91

Selected Scale Intercorrelations - Grade 5 Reading
GROUP B
N=65

		V2 (a)	V2 (b)	M1	M2	M3	M4	T1	T2	T3	T4	TG	T5	01,2	03	A	R	V
Mng-Derived	V2(a)	-																
Mng-Affix	V2(b)	.47	-															
Multi-mng	M1	.48	.41	(.85)														
Unfamiliar	M2	.30	.46	.53	(.72)													
Pronoun Ref.	M3	.46	.45	.66	.54	(.66)												
Figurative	M4	.44	.47	.51	.48	.47	(.74)											
Recall	T1	.26	.43	.56	.57	.43	.49	(.47)										
Translation	T2	.36	.50	.51	.59	.55	.43	.55	(.56)									
Integration	T3	.26	.47	.54	.56	.53	.58	.62	.53	(.55)								
Extrapolation	T4	.40	.44	.65	.56	.55	.53	.57	.70	.62	(.63)							
Main Idea	TG	.45	.46	.60	.47	.55	.55	.62	.67	.70	.77	-						
Evaluation	T5	.34	.47	.51	.43	.40	.32	.45	.44	.37	.51	.40	(.36)					
Topic	01,2	.54	.53	.65	.59	.67	.58	.39	.49	.48	.62	.42	.37	-				
Outline	03	.36	.24	.58	.39	.54	.28	.17	.32	.29	.36	.29	.37	.48	(.56)			
Applied	A	.41	.54	.47	.57	.64	.52	.39	.59	.46	.49	.47	.38	.62	.40	(.66)		
Ravens	R	.23	.25	.28	.34	.55	.30	.20	.39	.35	.36	.38	.18	.36	.29	.52	-	
Vocabulary	V	.54	.62	.67	.50	.60	.56	.54	.62	.45	.64	.52	.50	.61	.44	.56	.50	(.87)

Selected Scale Intercorrelations - Grade 2 Math
GROUP A
N=58

Scales		C1	C2	C3	C4	C5	C6 +7	C8	C9	C10	N1	N2	N3	P1 a	P1 b	P1 c	P1 d	R	V
Add. No Regrp.	C1	(.80)																	
Add. with Regrp.	C2	.72	(.87)																
Sub. No Regrp.	C3	.74	.62	(.92)															
Sub with Regrp.	C4	.49	.43	.60	(.90)														
Multiplication	C5	.40	.40	.21	.17	(.93)													
Speed Tests	C6 +7	.77	.59	.80	.59	.22	(.86)												
Inequalities	C8	.44	.41	.51	.60	.30	.53	(.76)											
Number Line	C9	.26	.17	.22	.09	.25	.22	.51	-										
Renaming Numerals	C10	.53	.50	.62	.59	.32	.68	.55	.20	(.81)									
Numberals	N1	.45	.44	.45	.24	.38	.37	.23	.45	.37	-								
Ordinals	N2	.60	.39	.51	.37	.42	.56	.40	.43	.37	.52	-							
Place Value	N3	.59	.48	.46	.44	.42	.67	.34	.23	.57	.34	.11	-						
Properties	P1 (a)	.24	.15	.35	.26	.42	.19	.31	.11	.24	.31	.23	.31	-					
	(b)	.20	.25	.26	.17	.38	.22	.25	.19	.22	.16	.13	.20	.57	-				
	(c)	.22	.29	.24	.28	.42	.15	.18	-.01	.16	.17	.19	.34	.70	.74	-			
	(d)	.18	.24	.14	.42	.17	.23	.29	.06	.25	-.06	-.07	.39	.16	.21	.24	-		
Ravens	R	.29	.28	.14	.20	.39	.05	.33	.32	.13	.40	.22	.23	.41	.24	.38	.11	-	
Vocabulary	V	.36	.14	.31	.22	.33	.40	.39	.40	.32	.40	.30	.52	.42	.21	.26	.17	.39	-

Selected Scale Intercorrelations - Grade 2 Math
GROUP B
N=74

		N4 (a)	N4 (b)	NS	M1	M2	M3	M4	M5	G1	G2	A1	R	V
Fractions: sets	N4 (a)	(.30)												
Fractions: regions	N4 (b)	.50	(.70)											
Odd & Even #'s	N5	.29	.26	(.70)										
Money	M1	.31	.28	.40	(.63)									
Linear	M2	-.02	.08	.32	.37	(.66)								
Gen. Meas.	M3	.20	.08	.12	.37	.25	-							
Liquid	M4	.21	.27	.19	.22	.26	.31	(.34)						
Time	M5	.24	.30	.26	.22	.24	.10	.17	(.62)					
Figures	G1	.31	.34	.22	.39	.37	.31	.28	.25	-				
Curves/points	G2	.44	.28	.18	.28	.30	.32	.20	.12	.39	-			
Word Problems	A1	.25	.41	.39	.39	.32	.17	.48	.31	.35	.43	(.61)		
Ravens	R	.17	.17	.38	.23	.25	.32	.27	.39	.26	-.001	.24	-	
Vocabulary	V	.26	.35	.39	.27	.25	.32	.33	.38	.46	.23	.40	.42	-

Selected Scale Intercorrelations - Grade 5 Math
GROUP A
N=72

		C1 (a)	C1 (b)	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	N1	N2	N4 (a)	N4 (b)	N4 (c)	N5 (a)	N5 (b)	N6 (c)	N6
Whole Number (+)	C1(a)	(-)																						
Whole Number(-)	(b)	.43	(-)																					
Whole Number (x)	C2	.24	.42	(.81)																				
Whole Number (÷)	C3	.22	.43	.57	(.72)																			
Fractions (+)	C4	-.02	.24	.22	.18	(.81)																		
Fractions (+)	C5	.05	.14	.21	.10	.77	(.79)																	
Fractions (+)	C6	-.06	.19	.32	.33	.52	.46	(.80)																
Fractions (-)	C7	.18	.32	.26	.29	.43	.36	.65	(.64)															
Fractions (-)	C8	-.01	.24	.28	.29	.57	.48	.68	.60	(.81)														
Fractions (x)	C9	-.06	.09	.17	.24	.27	.15	.42	.36	.60	(.60)													
Fractions (÷)	C10	-.18	-.08	-.07	.08	.01	-.08	.12	.12	.09	.31	(.36)												
Decimals (+, -)	C11	.37	.37	.57	.56	.15	.13	.25	.10	.09	-.01	-.07	(.85)											
Inequalities	C12	.13	.31	.36	.39	.24	-.03	.10	.13	.18	.24	-.03	.28	(.74)										
Speed Test	C13	.35	.45	.60	.59	.25	.10	.19	.22	.23	.17	.09	.52	.56	(.84)									
Place Value	N1	-.05	.02	.12	.18	.03	.08	.00	-.06	.12	.07	.02	.04	-.02	.19	(.52)								
Properties	N2	.11	.34	.28	.51	-.01	-.05	.12	.22	.09	.14	-.04	.19	.35	.44	.01	(-)							
Multiples	N4(a)	.03	-.02	.04	.02	-.25	-.33	-.10	.06	-.15	-.01	-.07	.13	.28	.12	-.11	.15	(.52)						
Multiples	(b)	-.16	.14	.17	.08	.04	.07	-.06	.00	.16	.03	.09	-.19	-.08	-.12	.09	-.05	.00	(-.02)					
Multiples	(c)	-.16	.09	-.17	.00	.09	.14	-.01	.00	.20	.04	.10	-.18	-.14	-.18	.05	-.18	.00	.94	(.54)				
Factors	N5(a)	.12	.11	-.16	.11	-.27	-.31	-.35	-.17	-.20	-.06	-.07	.04	.28	.19	-.08	.35	.32	.11	-.02	(.45)			
Factors	N5(b)	.08	-.02	.12	.13	-.20	-.17	-.25	-.09	-.02	-.01	.06	.10	.32	.19	-.03	.16	.23	.05	-.01	.30	(.34)		
Factors	(c)	.00	.00	.00	.08	-.26	-.23	-.24	-.19	-.01	.21	.21	.09	.17	.25	.08	.25	.32	.11	.07	.32	.41	(.53)	
Factorization	N6	.11	.14	.00	.12	-.19	-.13	-.25	-.12	-.21	-.02	.02	.03	.11	.14	-.03	.49	-.07	-.06	-.18	.33	.11	.29	(-
Ravens	R	.30	.32	.39	.31	.13	.14	.38	.33	.25	.02	-.16	.37	.33	.28	.08	.22	.13	-.25	-.21	-.01	.02	-.12	.02
Vocabulary	V	.45	.47	.41	.46	.16	.04	.23	.23	.30	.15	-.15	.41	.56	.55	.10	.43	.24	-.17	-.20	.17	.30	.25	.25

Selected Scale Intercorrelations - Grade 5 Math
GROUP B
N=68

		G1	G2	G3	G4	G5	G6	G7 (a)	G7 (b)	G8	S1	S2	S3	S4	A1	A2	A3	A4	A5	R	V
Perimeter	G1	(.34)																			
Area	G2	.16	(-)																		
Volume	G3	.07	.28	(-)																	
Surf. area	G4	.00	.29	.27	(-)																
Volume	G5	.18	.31	.07	.20	(-)															
Lines	G6	.16	.01	-.27	-.13	.12	(.60)														
Figures-Plane	G7(a)	.23	.05	.01	.10	.05	.41	(-)													
Figures-Solid	G7(b)	.34	.14	-.03	.10	.05	.25	.50	(-)												
Circle	G8	-.03	.16	.04	.03	-.06	.27	.38	.17	(.52)											
Number pairs	S1	.48	-.01	-.14	-.12	-.19	.07	.50	.11	-.22	(-)										
Graphs	S2	.50	.33	-.02	-.04	-.09	.19	.24	.13	.11	.58	(-)									
Average,etc.	S3	-.05	.04	.21	.00	-.19	.16	.17	.02	.26	-.07	.12	(.44)								
Number patterns	S4	.14	.09	-.02	.04	.14	.19	.42	.40	.11	.14	.35	.10	(.70)							
Word Prob. 1-step	A1	.16	.09	-.04	.03	.14	.34	.40	.46	.27	-.12		.14	.53	(.36)						
Word Prob. 2-step	A2	.25	.09	-.07	.27	.06	.13	.34	.29	.27	.20	.28	.12	.44	.43	(.60)					
Word Prob. Metric	A3	.22	-.02	.27	.22	-.08	-.04	-.04	.08	.12	-.03	-.05	.25	-.02	.12	.24	(-.02)				
Word Prob. Measures	A4	.34	.21	.22	.30	.17	.05	.19	.25	.23	-.03	.09	.14	.24	.41	.60	.25	(.54)			
Word Prob. Average	A5	.13	-.05	.04	.01	.12	.06	.08	.12	.09	-.05	.00	.10	.31	.28	.29	.21	.12	(.33)		
Ravens	R	.32	.22	-.16	.09	.12	.02	.11	.27	.09	.19	.30	-.14	.34	.26	.43	.05	.17	.24	(-)	
Vocabulary	V	.37	.11	-.08	.09	.02	.28	.40	.39	.32	.10	.38	.14	.42	.41	.61	.13	.35	.27	.46	(-)